SANDIA REPORT

SAND2013-0940
Unlimited Release
Printed February 2013

Development of a Statistically Based Access Delay Timeline Methodology

W. Gary Rivera, David G. Robinson, Greg D. Wyss, Stacey L. Hendrickson

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2013-0940
Unlimited Release
Printed February 2013

Development of a Statistically Based Access Delay Timeline Methodology

W. Gary Rivera
Active Response and Denial
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0783
wgriver@sandia.gov

David G. Robinson
Data Analysis and Informatics
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1327
drobin@sandia.gov

Gregory D. Wyss
Security Systems Analysis
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0757
gdwyss@sandia.gov

Stacey Langfitt Hendrickson
Human Factors and Statistics
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0830
smhendr@sandia.gov

Abstract

The charter for adversarial delay is to hinder access to critical resources through the use of physical systems increasing an adversary's task time. The traditional method for characterizing access delay has been a simple model focused on accumulating times required to complete each task with little regard to uncertainty, complexity, or decreased efficiency associated with multiple sequential tasks or stress. The delay associated with any given barrier or path is further discounted to worst-case, and often unrealistic, times based on a high-level adversary, resulting in a highly conservative calculation of total delay. This leads to delay systems that require significant funding and personnel resources in order to defend against the assumed threat, which for many sites and applications becomes cost prohibitive. A new methodology has been

developed that considers the uncertainties inherent in the problem to develop a realistic timeline distribution for a given adversary path. This new methodology incorporates advanced Bayesian statistical theory and methodologies, taking into account small sample size, expert judgment, human factors and threat uncertainty. The result is an algorithm that can calculate a probability distribution function of delay times directly related to system risk. Through further analysis, the access delay analyst or end user can use the results in making informed decisions while weighing benefits against risks, ultimately resulting in greater system effectiveness with lower cost.

Acknowledgment

This research team would like to recognize and show our gratitude for the contributions of the Laboratory Directed Research and Development Program at Sandia National Laboratories for the opportunity to conduct this investigation.

The team would also like to thank Brad Parks (Organization 6630) and Chad Monthan (Organization 6634) for their support and input throughout the proposal, investigation, and reporting of this work.

Contents

Preface **14**

Summary **16**

Nomenclature **16**

1 Introduction **17**

1.1 Problem Definition 17

1.2 Background 18

1.3 Scope 19

1.4 Long Term Issues and Obstacles 20

2 Background **21**

2.1 Problem Description 21

2.2 Evaluation of Individual Attack Paths 23

2.2.1 Manual Evaluation of Facility Security 23

2.2.2 Evaluation With Uncertain Task Durations 24

2.3 Automated Evaluation using Point Estimate Data 26

2.4 Other Directed Graph Approaches 29

2.4.1 Multiple Attack Paths 29

Stochastic Shortest Path 29

Deterministic Edge Weights 30

Random Edge Weights with Known Distributions 31

2.4.2 Monte Carlo Discrete Event Simulation 32

2.5 Need for a New Approach 33

3 Introduction to Bayesian Methods **35**

3.1 Background 35

3.2 An Alternative Approach 35

3.3 Basic Concepts 37

3.4 Example 39

3.5 Risk and Benefits of the New Approach 39

3.5.1 Risk: Establishing Priors 39

3.5.2 Benefits 39

4 A New Approach **43**

4.1 Introduction 43

4.2 Case I: Intruder Detected Immediately 44

4.3 Case II: Individual Detection Events 45

4.4 Case III: Intrusion Failure Due to Task Failure 47

5 Integrating Human Factors **49**

5.1 Background 49

5.2 Workload 49

5.3 Ergonomics 50

5.4 Fitness for Duty 52

5.5 Time Pressure 52

5.6 Accounting for Effect 53

5.7 Adversary Skill Levels 54

6 Demonstration of Approach **57**

6.1 Problem Description 57

6.2 Traditional Access Delay Calculation 57
6.3 Statistical Analysis 60
6.3.1 Task Times 60
6.3.2 Construction of Priors 61
6.4 Results 62
6.5 Implementation Summary 69
6.5.1 Characterize Adversary Paths 69
6.5.2 Collect Historical Data 69
6.5.3 Conduct Testing 70
6.5.4 Characterize Adversary Path Risk 70

7 Conclusions and Recommendations **71**
7.1 Conclusions 71
7.2 Future Work 76

References **78**

Appendix

A Dijkstra Example **83**
A.1 Dijkstra's Algorithm 83

B Introduction to Bayes Theorem **87**
B.1 Impact of Testing 87
B.2 Second Example 88
B.2.1 Black Swan 89
B.2.2 Bayesian Approach 89
B.2.2.1 Prior 89

B.2.2.2 Posterior 90

C Example - Case I: Intruder Detected Immediately **93**

C.1 Background 93

C.2 Results 94

D Example - Case II: Individual Detection Events **97**

D.1 Uncertainty Foundation 97

D.2 Scenario 99

D.2.1 Demonstration Results 102

E Test Data **107**

List of Figures

2.1 Representation of adversary tasks for a theft attack. An access attack would be complete with Task 6, and a sabotage attack would add a different Task 7 to represent the time required to complete the sabotage action itself. Task 8 represents the time required to escape from the facility. 21

2.2 Representation of the EASI model as an event tree for an example with three detection opportunities. 26

2.3 Adversary Sequence Diagram (ASD) for a sample facility. Functional protection elements form layers of protection around the facility. Note that Room Surface 2 is a shared wall between the Controlled Building Area and the Target Enclosure, so the adversary is able to create a direct path and jump between these two layers without entering the Controlled Room. [16] 28

2.4 Multiple Attack Paths 30

2.5 Typical Directed Acyclic Graph 31

3.1 Flow of Information in Bayesian Analysis 38

3.2 Impact of Combining Test Data with Prior 40

3.3 Sample Priors Based on Intruder Type 41

4.1 Stochastic Shortest Path w/Random Edge Weights 44

4.2 Case I: Basic Description 45

4.3 Case 1 Distribution Interference 45

5.1 Performance versus arousal/perceived stress curve [11] 50

5.2 Priors Based on Intruder Type 55

6.1 Facility Barrier-Task Decomposition 60

6.2 Prior on $\mu_z = E[ln(t)]$ 62

6.3 Prior on $\tau_z = 1/V[ln(t)]$ 63

6.4 CDF for Barrier Breach Times (t-seconds) 64

6.5 CDF - Facility Breach Time (t- seconds) 65

6.6 Pr{Response Team Intercepts Adversary} (t- seconds) 66

7.1 Delay as a Function of Median Time 73

7.2 Delay as a Function of Minimum System Response 74

7.3 Access Delay For Adversary Skill Level 75

A.1 Populated DAG for Dijkstra Algorithm 84

C.1 Case 1: PDF of Task Times 95

C.2 Case 1: Total Breach Time 95

C.3 Case 1: Response Time 95

C.4 Case 1: Delta Time (Delta $> 0 =$ success) 96

D.1 Demo Problem: Stochastic Breach, Response Times, and Detection Probability . . . 97

D.2 Demonstration System 97

D.3 Prior Distribution on Detection Probability Parameters 100

D.4 Prior Distribution on Detection Probability 101

D.5 Prior Distributions on α_i, λ_i 102

D.6 Probability of Intruder Interception P_s 103

D.7 Probability of Detection P_D 103

D.8 Summary of Important Random Variables 104

D.9 MCMC Convergence Check - Gelman Statistic 105

List of Tables

6.1 Path Time for Average 58

6.2 Path Time for - 50% 59

6.3 Barrier 1 Data 61

6.4 Individual Task Times - Quantiles 67

6.5 Barrier Breach Times - Quantiles 67

6.6 Comparison of Traditional v Statistical 68

A.1 Steps in Dijkstra Algorithm 85

C.1 Case 1: Distributions for Generating Simulated Data 94

C.2 Case 1: Results 94

D.1 Data Table 99

D.2 Results From Initial Demonstration Problem 103

E.1 Barrier 1 Data 107

E.2 Barrier 2 Data 107

E.3 Barrier 3 Data 107

E.4 Barrier 4 Data 108

E.5 Barrier 5 Data 108

E.6 Barrier 6 Data 108

E.7 Barrier 7 Data 109

Executive Summary

Traditional access delay methodology involves adding up a series of delay times for a series of obstacles in an adversarys path to determine total delay for a system. These obstacles can be as complex as sophisticated barrier systems requiring significant skill and equipment to breach, or as simple as a distance over that an adversary must traverse to get to the next barrier or target. Measured delay and task rate data is often sparse and is rarely considered statistically viable in the classical sense. Further, for any task conducted by different people, tools, or techniques, there will be a range in the time required to complete a task or series of tasks to overwhelm a barrier, and ultimately complete the path.

Classical statistical methods, those based on a frequency approach, permit the combination of point estimates but stumble when confidence in system level performance must be characterized or when minimal data is available. Alternatively, the use of a Bayesian methodology provides a formal process for synthesizing and learning from the data that is in agreement with the current information sciences. It permits relevant expert opinion, sparse field data, and laboratory test information to be objectively codified and integrated.

This report describes a new methodology that considers the uncertainties inherent in the problem to develop a realistic timeline distribution for a given adversary path, taking into account the uncertainty in each of the model parameters. This methodology calculates adversary delay by advanced Bayesian statistical methodologies, taking into account small test sample size, expert judgment, and defined threats.

Because this has been a starting point, this investigation has focused first on the mathematical structures to handle small data set problems, second on the steps to mathematically combine multiple sequential tasks to form an overall delay distribution for a given scenario and set of data, and finally to present an algorithm that could be used to conduct actual delay analysis. The analysis approach is outlined for the situation where the breach of each barrier may involve multiple tasks, each of which must be accomplished before the barrier is successfully breached.

Human performance factors must be considered when seeking to characterize any access delay system. Several variables may impact human performance on the type of tasks needing to be completed by an adversary during an attack. Fatigue, workload, lighting, noise, extreme temperature, and task switching are some of the stressors that can affect a persons performance. Applying multipliers to the developed calculations is discussed.

This research has shown that by using Bayesian techniques, mathematically defendable probability density functions representing barrier breaching times can be generated for small data sets. Using these calculated time distributions, a sequential string of tasks/barriers (time distributions) can be combined to develop a probability density function for the time required to complete the

chosen path.

Nomenclature

ASD Adversary Sequence Diagram

CDP Critical Detection Point

DBT Design Basis Threat

EASI Estimate of Adversary Sequence Interruption

HEP Human Error Probability

HRA Human Reliability Analysis

PIF Performance Influencing Factors

PPS Physical Protection System

SME Subject Matter Expert

P_C Probability that the need for the response force is successfully communicated

P_D Probability of detection

P_E Probability of effectiveness

P_N Probability of neutralization

P_{ND} Probability of non-detection

P_{NI} Probability that scenario is not interrupted in time

T_k Normal distribution of task time for each remaining adversary task (EASI model)

T_R Response force time

Shortest Path Analysis Dominant current practice. The shortest path from entry to target is found using, for example, worst case delay times.

Stochastic Shortest Path Delay and task times are now random variables. Assume that no response is available.

Chapter 1

Introduction

1.1 Problem Definition

The charter for adversarial delay is to hinder access to critical resources through the use of physical systems to increase an adversary's task time. Assuming prompt adversary detection, the result is an increased time for response forces to mobilize, arrive, interrupt, and neutralize the adversary.

The traditional method for characterizing access delay has been a simple model focused on accumulating times required to complete each task with little regard to uncertainty in the model, model parameters, task complexity, or decreased efficiency associated with multiple sequential tasks or stress.

The delay associated with any given barrier or path is often further discounted to report worst case, and often unrealistic, delay times based on a high-level adversary team. The result is a conservative calculation of total delay where all attack paths are considered viable in the minimum possible time. This methodology results in the requirement for high levels of funding and personnel resources to defend attack scenarios which would have little chance of adversay success with reasonable adversary timelines.

Complicating the analysis is that measured delay and task rate data is often sparse and is rarely considered statistically viable in the classical sense. Budgetary and schedule constraints often limit testing to data sets with six or fewer repetitions, with many tests only able to conduct one, two, or three repetitions for any given test scenario or task. In some cases subject matter expert (SME) opinion and/or subscale testing are used to fill in known data deficiencies.

This report proposes a new methodology that considers the uncertainties inherent in the problem to develop a realistic timeline distribution for a given adversary path, taking into account the uncertainty in each of the model parameters. By specifically considering minimal test data that takes into account adversary ability, tool set, breach time distribution, probability of successfully breaching a barrier, as well as barrier effectiveness to these factors, a model can be developed to represent the effectiveness of the delay system against various attack scenarios and threats.

This methodology calculates adversary delay by novel statistical methodologies, taking into account small test sample size, expert judgment, and defined threats. The goal has been to create an easy to communicate distribution of delay times for a given path in terms of adversary threat.

Armed with this information, system designers and security personnel can then focus their limited resources on defeating attack scenarios that are most advantageous to an adversary, ultimately resulting greater system effectiveness with lower cost.

1.2 Background

Traditional access delay methodology involves simply adding up a series of delay times for a series of obstacles in an adversarys path. These obstacles can be as complex as sophisticated barrier systems requiring significant skill and equipment to breach, or as simple as a distance over flat ground that an adversary must traverse in order to get to the next barrier or target. The time required to breach any given barrier can be further broken into smaller or individual tasks that will comprise the overall delay for the barrier or barrier system. Once a list of tasks are agreed upon, the time required to complete each task is assigned, and the final reported delay for that path is given as the sum of the task times required to complete the path.

It should be noted that task time is not considered *delay* unless detection has occurred, because only when the adversary is detected does the response force know that it must assess the detection and respond accordingly.

As with any task conducted by different people, different tools, or different techniques, there will be a range in the time required to complete a task or series of tasks to overcome a barrier, and ultimately complete the path. The potential factors that affect the delay time are numerous and are not always easy to characterize. As a result when testing is conducted to determine delay time through a barrier, individual task time (single test), average times, and minimum times are recorded and used in different ways. Because it is possible, depending on the requirements of the analysis, to use minimum times, average times, or some hybrid calculation of minimum, average, and maximum, different path delay times can be calculated for the same set of steps. Currently, statistical methods are rarely used to manually calculate task times, and most often are simply calculated by adding comparable elements (minimum, average, or less often maximum). One methodology, for instance, is to calculate the average time to breach a barrier, then subtract 50% of that value for the assumed minimum time and add 50% of the average to give the assumed maximum time (effectively resulting in a symmetric triangular distribution). Negative 50% of the average is chosen to account for potentially high level adversaries, while adding 50% of average simulates a less capable adversary.

Further complicating current analysis is the use of *delay multiplication factors*. These are delay mechanisms such as combat stress, being put under small arms fire, visual obscurants (which are not technically a barrier), sensory irritants (e.g. loud sounds), etc. Typically, it is left to the analyst to apply delay multiplication factors based, in some cases, on test data or SME judgement.

The problem of developing a methodology that is generally applicable to a wide variety of facilities and attack scenarios is extremely complex with a plethora of dark corners for in-depth investigation. The intent of this effort is to develop a solid foundation for future efforts. The goal

has been to develop a new approach and demonstrate the potential through the application and testing on a sample problem with real, measured data.

1.3 Scope

The scope of this research effort is primarily to investigate statistical methodologies that may be better applicable to the realities of practical access delay analysis.

Because this is a starting point, this investigation will focus first on the mathematical structures to handle small data set problems, second on the steps to mathematically combine multiple sequential tasks to form an overall delay distribution for a given scenario and set of data, and finally to propose an algorithm that could be used to conduct actual delay analysis. Specifically out of scope for this effort is:

- Probability of Neutralization (P_N) of the Adversary: This research focuses on determination of adversary task/path time (which can also be applied to response force task/path time by simply redefining tasks). The requirement for the response force to neutralize the adversary force once the adversary is interrupted is beyond the scope of this project.

- Adversary Pathing Analysis: To date, a significant amount of research has been undertaken by others at SNL and elsewhere that looks into adversary pathing strategies. This analysis often takes into account fastest path, path of least detection, adversary tool set, and even insider information. This project does not seek to characterize the statistical likelihood of an adversarys choice of path. Instead, the current work relies on an analyst choosing the most appropriate path for analysis. That said, the methods discussed in the following chapters could be incorporated into a larger pathing analysis.

- A User-Ready Access Delay Tool: The present research investigates the applicability of new and novel statistical methodologies up to the point that an access delay calculation can be accomplished. But this will likely involve statistical and computer tools outside of the common operating environment of most delay analysts. The ultimate end goal of this work is to develop a computer tool that an analyst can use in the field by inputting a set of required data, along with a pre-programmed database of access delay task times and adversary capabilities. This final step will require additional funding from a source yet to be named.

- A level determination of adversary capability: Recently, the difference in delay times associated with different levels of adversary capability has been discussed. Currently, the precise definitions of high, medium, and low level adversaries are still in process. This research does not propose to define adversary task time as it relates to capability. That said, the statistical methodologies proposed can be applied to any level of adversary given a viable data set. Further, a discussion on how the methodology output can be applied to the different adversary level categories.

The following presented research is focused on the approach for the general problem of calculating a distribution of delay times for a single path or set of obstacles. The discussion begins with a very broad perspective and outlines in specific terms, the mathematics required to tackle the complete problem. The result is a framework for the solution to the larger problem, with the understanding that it is not possible to tackle something of this complexity under the current research project. Rather, the larger solution highlights the general requirements for the foundation to be laid by this project. It is critical to understand what hooks need to be in place so that the results of the current effort can be extended to tackle the larger site security problem.

The discussion builds from the perspective of deterministic breach times using Dijkstras Algorithm as an example of one approach. The discussion is then extended to stochastic breach times with known probability distributions and concludes with the most general case where the distribution of breach times is not known. This latter situation is a necessary goal given that there can be limited data available on various security aspects, e.g. breach time, and it might be required to merge data from expert opinion and various field testing.

1.4 Long Term Issues and Obstacles

The long term problem to be addressed is highly complex, and there are a number of major characteristics of the problem that need to be addressed for a complete answer:

- Multiple intruders or intruder teams (i.e. coordinated attacks with parallel tasks)
- Single/multiple targets
- Multiple intruder paths (decision points)
- Uncertain delay/breach times
- Security monitors with detection error
- Uncertainty in response time
- Limited test data, possibly from multiple sources

A concerted effort was made to assure that the methodology developed had the capability to be extended to account for the above issues. First order assumptions are to limit the analysis to a single group of intruders and a single target. As a minimum, the analysis must account for random delay times and random response times. Allowances will be made for possible extension to multiple paths with random delay times.

Chapter 2

Background

2.1 Problem Description

The objective of our methodology development activity is to analyze the performance of a physical protection system for a facility. A physical protection system (PPS) performs well or is *effective* to the degree that it prevents adversaries from achieving their desired objectives. A typical metric for effectiveness is the probability P_E that the PPS successfully prevents a prototypic adversary (a *design basis threat*, or DBT) from accomplishing a successful attack. Adversaries may want to steal an item, damage an item, damage a facility, collect information, or any of a number of other tasks down to simply generating publicity by gaining access to a restricted area or even just attempting an attack. An attack usually requires the adversary to successfully complete a series of tasks as shown in Figure 2.1.

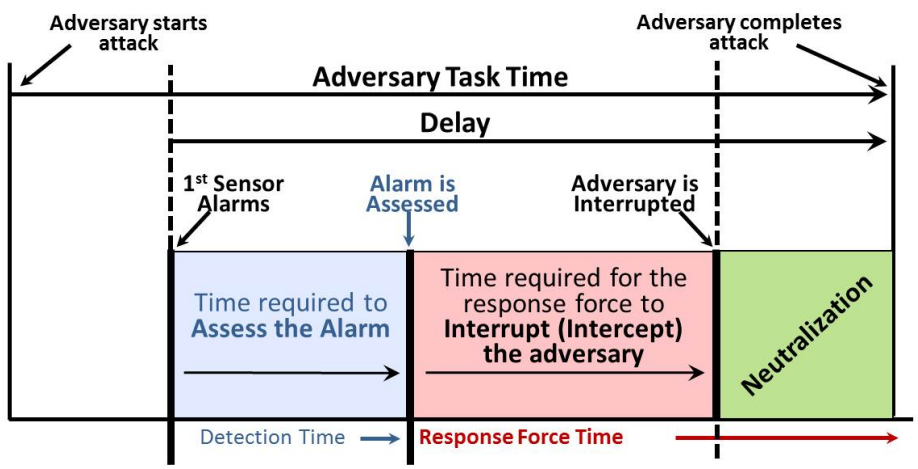

Figure 2.1. Representation of adversary tasks for a theft attack. An access attack would be complete with Task 6, and a sabotage attack would add a different Task 7 to represent the time required to complete the sabotage action itself. Task 8 represents the time required to escape from the facility.

In order for the adversary to be successful, he must accomplish the desired task before the defenders can respond to interrupt the task and neutralize the attack. The task might involve a theft, in which the adversary must travel from a starting point that is usually outside the facility to the target item inside the facility, obtain the desired item, and travel with the item to the outside of the facility. If the task is to gain access or cause damage to a facility, the adversary need only make the first part of the journey, encumbered by any tools needed to gain access or to commit the desired action.

The physical security problem, as described above, is fundamentally a race. For the defenders to be successful, they must detect that they are being attacked and delay the adversary from completing their objective before defenders can respond to interrupt the task and neutralize the attack. Key data used to evaluate the PPS include the probability that the defenders will detect the adversarys attack during each attack task (which typically results in dispatch of a response force to interrupt and stop the attack), the time taken by the adversary in performing each attack task, and the time taken by the response force to arrive with sufficient force to interrupt the attack once the force is dispatched. The adversaries are *interrupted* when they must stop their tasks and deal with the response force, and are *neutralized* when they no longer pose a threat. The security mantra *detect, delay, respond* captures these major features of the PPS, and encapsulates the wisdom that one must detect the attack first, as an undetected adversary can (theoretically) spend as much time as desired on each attack step because the response force will not be dispatched to even attempt to interrupt the attack tasks.

This problem formulation sets up a race between the adversary and defender. Once detected, the adversary must complete the attack tasks before the response force can arrive and interrupt them. Thus, the adversary wins if the response force interruption time T_R is greater than the sum of the times for the attack tasks that occur after the attack is detected. Furthermore, an adversary might decide to manage the point at which detection would be expected to occur by using stealth or deceit for the early stages of the attack when detection is not desired, then transition into a high-speed overt attack for those final attack tasks that can be accomplished before the response force arrives. This shows that early detection is often a key to defender success. Analyses often examine an attack scenario to determine the last point at which detection of the adversary will lead to a response that is quick enough to prevent the adversary from successfully completing the attack. This point is known as the Critical Detection Point (CDP) for the attack scenario. The adversary's optimal attack, then, may be to minimize their probability of being detected up to the CDP, then minimize task time once beyond the CDP. If one assumes that the response force is highly likely to respond effectively within the time T_R, then the probability that the attack scenario will be interrupted P_I is simply the aggregated probability of detection over all attack steps prior to the CDP. Once the response force arrives to interrupt the attack, the responders must finish the task by neutralizing the adversary (often thought of as *winning the gun battle*). The probability that the response force can neutralize the adversary given that interruption has occurred is P_N, but neutralization cannot occur unless the attack has been successfully interrupted. The PPS is successful if interruption and neutralization both occur, so $P_E = P_I P_N$. Estimation of P_N is beyond the scope of this report, but involves a variety of methods such as simulated combat. The purpose of this project has been to develop a more statistically sound method for estimating P_I.

Numerous tools have been developed to help in the calculation of P_E and its components. The tools that are used for estimation of P_I will be discussed in the following sections with their assumptions and limitations. This leads to a discussion of the need for a new method that can relax certain assumptions and eliminate many of the limitations of the reviewed methods.

2.2 Evaluation of Individual Attack Paths

2.2.1 Manual Evaluation of Facility Security

The simplest method for evaluating P_I is a manual algorithm that is taught in a variety of courses and described in [16]. The evaluation can be done using pencil and paper, and requires only basic math. It is a simple instantiation of the framework described in the previous section, and assumes that an attack is attempted by a single adversary or a single group of adversaries who travel together along a single pathway as they make their way to the target and, if necessary, back out of the facility. This method does not discover attack pathways from a facility model, but rather evaluates a single attack pathway that is of interest to the analyst.

The method begins with a simple diagram of the attack path much like the one shown in Figure 2.1. The adversary tasks are identified, and the data for each task are listed in a table, including a description of the task and its required tools, the expected adversary time to complete the task, and the probability that the adversary will be detected while carrying out the task. The time and detection probability will naturally vary depending on the specific method used by the adversary in the task. For example, a cinder block wall can be breached with explosives or a jackhammer or hand tools. The hand tools may take longer than the explosives or jackhammer, but one might imagine circumstances where hand tools could be used with a much lower probability of detection than the other tools, if the adversary were to find that desirable. Thus, the task time and detection probability must be specific to the particular task and tools, and for this evaluation method, a single point estimate value is used for every time and probability considered.

The calculation begins with the analyst considering the T_R and beginning to accumulate task time starting at the end of the attack scenario (the final escape for a theft, or access to the target for many other types of attacks). Task by task, the attack times are added in order until the total the *remaining task time* just exceeds T_R, which identifies the CDP. If the attack is detected at or before this point, the response force can interrupt the attack, so the detection probabilities are aggregated for all earlier tasks in the attack. These probabilities are assumed to be independent (unless extenuating circumstances or common cause failure events might indicate otherwise). Since, to be successful, an adversary must avoid detection during all of these tasks, we compute the probability that detection fails at all relevant detection opportunities by multiplying the nondetection probabilities together ($P_{ND} = 1 - P_D$) to obtain the overall nondetection probability for the attack scenario P_{NI}, which is the probability that the scenario is not interrupted in time for the defenders to mount an effective response. P_I is computed as $(1 - P_{NI})$, and the computation is complete. If alternative attack paths are to be evaluated, the process is repeated, and the attack with the lowest

P_I is determined to be most advantageous to the adversary as it gives them the highest likelihood of successfully completing the attack without being interrupted or neutralized.

Because of the simplicity of the calculation, point estimate values (estimates without uncertainty) are used in these calculations. In order to ensure that a facility that is evaluated using this method is not wrongly characterized as adequately secure, the times and probabilities used in the calculation are deliberately chosen to be conservative. This means that the adversary task times are often selected to be near the low end of the observed data, the response time is selected to be near the high end of the observed data, and the detection probabilities are selected to be near the low end of the observed data. While this produces a conservative estimate for P_I, there have long been concerns that the repeated compounding of conservatisms within the calculation produces unrealistically short adversary timelines (the adversary finishes every task unusually quickly) and, therefore, an unrealistically early CDP and an unrealistically low value for P_I. In addition, the level of conservatism increases with the number of adversary tasks in the attack scenario. This can manifest itself in two unexpected ways. First, the longest and most complex attack scenarios often have the most unrealistically short timelines as they aggregate a larger set of conservative values than the simpler attack scenarios. Second, it is possible to obtain different adversary task timelines for the same attack scenario depending on the granularity with which the attack tasks are represented. Consider the task of using explosives to breach a wall. This task consists of several subtasks such as setting up the explosives, retreating to a safe distance, detonating the explosives, waiting for debris to clear, approaching the hole, and clearing debris from the hole. A conservative value for the single *breach wall using explosives* task might be much longer than the aggregation of conservative values for each of the six listed subtasks. Thus, the analyst must use care when selecting data and task granularity using this method.

This manual calculation method is surprisingly powerful, however, in that it is not limited to attack scenarios involving single targets. Consider a sabotage scenario in which an adversary must visit three rooms in order to access all of the equipment necessary to successfully complete the attack. This method simply requires the analyst to postulate the list and order of the attack tasks, and P_I can be computed regardless of the number of targets within the scenario. This is not true for many of the automated attack path search tools that will be discussed later, although it is important to note that for this manual calculation, the task of identifying the attack paths to be evaluated falls completely on the analyst.

2.2.2 Evaluation With Uncertain Task Durations

Whereas the manual calculation method estimates P_I for a single attack scenario only on the basis of point estimate data, a method that is mathematically more sophisticated has been developed to estimate P_I while considering uncertainties in the task times for both the attacker and defender. EASI (Estimate of Adversary Sequence Interruption) [6] is a simple method for evaluating the performance of a PPS that can be exercised using nothing more than a calculator with statistical functions. Like the manual calculation method, EASI does not discover attack pathways from a facility model, but rather evaluates a single attack pathway that is of interest to the analyst, and assumes that an attack is attempted by a single adversary or a single group of adversaries

who travel together along that single attack pathway. And like the manual calculation method, each detection opportunity is a point estimate value that represents the aggregate performance of the detector sensing abnormal or unauthorized activities, the transmission of the detector's signal to an alarm assessment point, and the realization by the human in assessing the alarm that an adversary attack requiring activation of the response force is in progress. However, unlike the manual calculation method, the duration of each adversary task in the attack scenario and the time required for the response force to react and interrupt the adversary are represented by normal distributions, each characterized by a mean and standard deviation. In addition, EASI introduces a probability, P_C, that the need for the response force is successfully communicated to initiate the response. The method by which the EASI model computes P_I is as follows. Each possible detection point is examined separately as though adversary detection occurs exactly at that point. The adversary is detected exactly at that point only if they are not detected at all previous detection points and are successfully detected at that point. This is computed by multiplying P_{ND} for all previous detection opportunities by P_D for the current opportunity. Now, given that the adversary is detected, interruption requires that the alarm be successfully communicated to the response force and that the remaining adversary task time is greater than the response force time T_R. The method convolves the normal distribution for the task time T_k for all remaining adversary tasks with the response force reaction time T_R to determine the probability that the response force arrives in time to interrupt the adversary. Thus, the P_I given detection at exactly the i-th detection opportunity P_{Ii} can be written as:

$$P_{Ii} = \left[\prod_{j=1}^{i-1}(1-P_{D_k})\right] P_{D_i} \cdot P_C \cdot Pr\left\{\left(T_R - \sum_{k=i}^{n} T_k\right) < 0\right\}$$

Since these computed interruption probabilities satisfy probabilistic conditions for being mutually exclusive, the overall P_I for this path is simply the sum of the interruption probabilities for each detection opportunity, or:

$$P_I = \sum_{i=1}^{n}\left(\left[\prod_{j=1}^{i-1}(1-P_{D_k})\right] P_{D_i} \cdot P_C \cdot Pr\left\{\left(T_R - \sum_{k=i}^{n} T_k\right) < 0\right\}\right)$$

It is also possible to view the EASI method graphically in terms of a small event tree. Each detection opportunity represents an event (or *question*) in the event tree that must be answered probabilistically based on the probability of detection or nondetection. At the end of the event tree one adds two events to represent the performance of the response force: one to represent the probability that the alarm is successfully communicated to the response force, and one to represent the probability that the response force arrives in time to interrupt the adversary. This probability will vary between paths within the event tree to represent the differences in remaining adversary task times as described previously. A representative event tree for a system with three possible detection opportunities is shown in Figure 2.2.

Note that the concept of a CDP is less clear when the task times are uncertain. One could theoretically compute the probability that each attack task is in fact the CDP through a series of

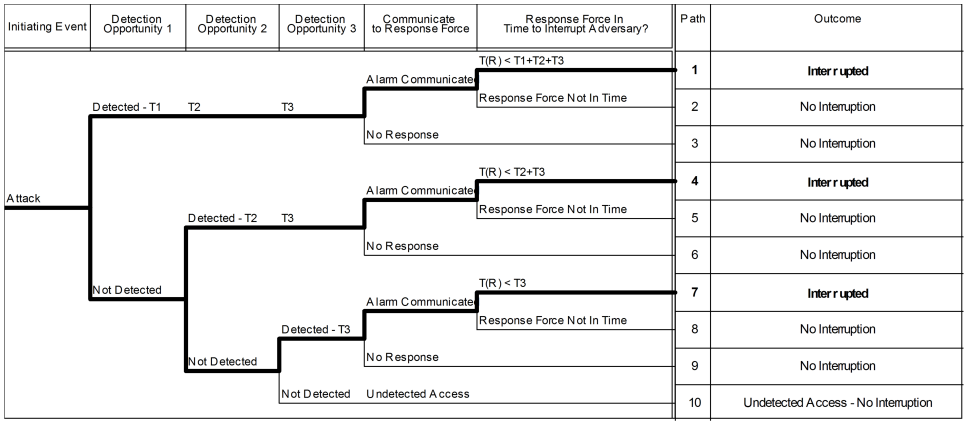

Figure 2.2. Representation of the EASI model as an event tree for an example with three detection opportunities.

additional convolution computations, although this is not done in EASI. The main assumptions and limitations for EASI are similar to those for the manual computation method described in the previous section, with the exception that all task and response times are represented as Normal distributions. EASI, as originally developed, limits the times to Normal distributions and cannot be used with other distribution types, although modern computational convolution methods can relax this limitation and allow the use of arbitrary distribution types. Nothing has been published in the literature that would guide an analyst through the selection of appropriate distributions or the computational process, and no work is available to quantify the benefits that can be realized by using other distribution forms. One important limiation relates to the fact that the Normal distribution used in EASI has infinite support. Thus, for any task, there is a mathematical artifact that it is *possible* to accomplish a task in less than zero time, which is not realistic. Use of distributions with limited support would eliminate this artificiality from the method.

2.3 Automated Evaluation using Point Estimate Data

Both of the methods described in the preceding sections are subject to one important limitation: they are applied for a single adversary pathway. If an analyst is to evaluate multiple pathways, they must be discovered through a separate analysis. Over the last few decades, software has been developed at Sandia National Laboratories that automates the process of exploring the model of a security system in order to discover the most advantageous adversary pathway using the manual calculation method described previously. The first generation software developed for this purpose is Analytical System and Software for Evaluating Safeguards and Security (ASSESS) [22], while the second generation software is Adversary Timeline Line Analysis Software (ATLAS) [1]. Both tools operate on a similar principle.

The analyst begins the adversary timeline analysis by constructing an Adversary Sequence Diagram (ASD). The ASD shows how an adversary force encounters various barriers and tasks as they move through a facility. Since it is common to design security systems under the concept of *layers* or rings of security, the ASD embraces this concept. The analyst starts at the target that would be sought by the adversary and works outward to the first layer of security, which would be the last barrier the adversary would encounter before accessing the target. For example, this might be the wall of the vault that contains the target. The analyst then examines this layer of security to identify the various opportunities that an adversary might have to overcome this barrier. For a vault, one might be able to break in through a wall or the ceiling (which might have different construction characteristics), or enter through a door (which could be could be in different states such as open, closed but unlocked, or closed and locked). The analyst then moves to the next layer of security (say, the outer wall of the building that contains the vault) and specifies the adversary opportunities for that barrier. The process is repeated until all layers of security are specified, i.e., the analyst has specified the first barrier that an adversary would encounter when arriving at the facility from uncontrolled space external to the facility. Note that there are situations in which an adversary might be able to bypass one or more security layers, such as by parachuting into an inner area of a facility, or when the vault wall is also the outside wall of the building. In addition, the characteristics of a barrier in the egress pathway for a theft scenario may be different from the characteristics of the same barrier during the entry phase of the attack because, for example, the barrier was destroyed during entry or the barrier has one-way characteristics (e.g., a crash bar on an emergency exit). These conditions must be accounted for in the ASD and when the model is solved by the software. An example ASD is shown in Figure 2.3.

Once the ASD is created, the software takes over and performs the analysis. First, each opportunity for an adversary to penetrate a layer, or *security element*, is characterized according to its many possible defeat methods. For example, a locked door can be penetrated by an adversary by using the key to unlock it, by picking the lock, by manipulating the latch mechanism, by kicking it in, by breaking the window (if present), by cutting a hole in it with a saw, by using explosives, or by any of a myriad of other methods. Each defeat method has its own characteristic task time, detection probability, and required adversary training and tool kit. The software has a library of these defeat methods and their characteristics that has been developed and validated over the course of many years, and can simply associate the appropriate characteristics with each barrier as it is described by the analyst. Thus, the analyst specifies the type of barrier that is present at each point in the ASD, and the software provides appropriate defeat methods and their characteristics from the validated library.

The ASD and the associated defeat methods for all barriers represent a series of choices that are available to an adversary. In reality, an adversary can choose among all of the system elements and defeat methods to gain access through the first security layer, then again choose among all of the system elements and defeat methods to gain access through the second security layer, and repeat this choice process until the target is accessed (and, for theft scenarios, until the adversary has successfully egressed from the facility with the target item). Each combination of system elements and defeat methods represents a unique attack path, and the software can automatically identify each path and compute a value of P_I to assess its attractiveness to an adversary. Mathematically, this problem is combinatorial, and the search space expands rapidly with the number of security layers,

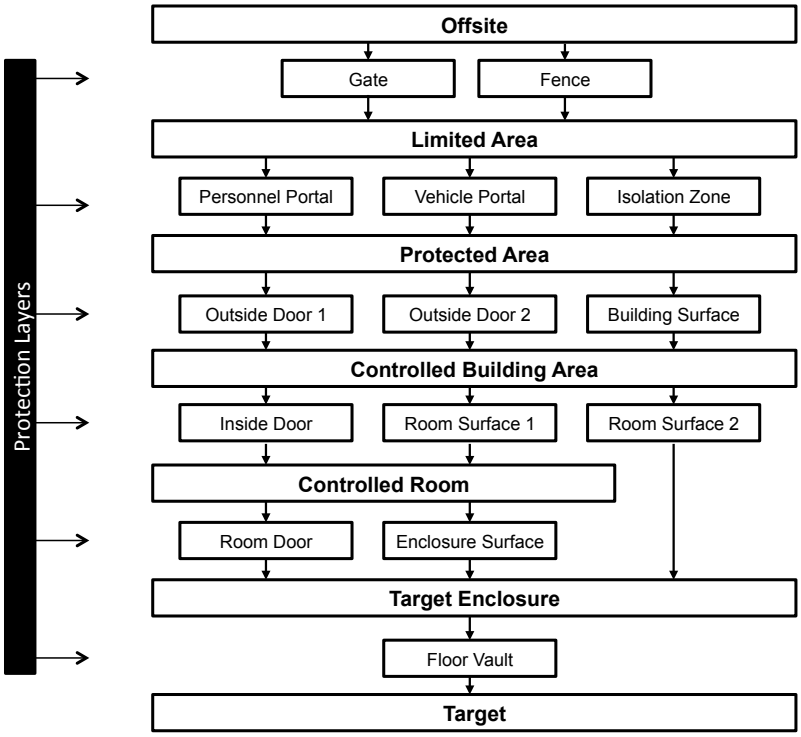

Figure 2.3. Adversary Sequence Diagram (ASD) for a sample facility. Functional protection elements form layers of protection around the facility. Note that Room Surface 2 is a shared wall between the Controlled Building Area and the Target Enclosure, so the adversary is able to create a direct path and jump between these two layers without entering the Controlled Room. [16]

system elements and defeat methods. In order to make the analysis computationally tractable, the search method is optimized using an efficient shortest path algorithm such as the Dykstra algorithm (see Appendix A). At the conclusion of the search, the software presents the analyst with the n attack paths that would be most desirable to an adversary because of their low P_I scores. Recall that a low value for P_I indicates that there is a low probability that the adversary would be interrupted before successfully completing the tasks of the attack.

The advantage of using an automated tool such as ATLAS or ASSESS is that it enables the analyst to sift quickly through the myriad possible adversary attack paths and understand which paths might be most attractive for an adversary and, thus, which paths should be subjected to more detailed scrutiny and possible mitigation or remediation. However, these tools have all of the remaining disadvantages of the manual calculation method in that they use only point estimate values for the protection force response time, adversary delay times and detection probabilities. Thus, the point estimate values used in the analysis are typically conservative (often to an unknown degree) and the conservatisms are compounded upon one another as longer attacks (attacks with more steps) are evaluated. While the algorithm seems to provide a reasonable rank ordering of the attractiveness of paths of comparable length, the analyst must treat the actual P_I values computed with significant skepticism until a further, more detailed analysis of the most attractive scenarios can be performed.

2.4 Other Directed Graph Approaches

The following discussion outlines some of the major approaches and issues for the most general case of using directed graphs to evaluate multiple attack paths.

2.4.1 Multiple Attack Paths

We will begin with the most basic generalization by the attack problem where multiple paths are possible (Figure 2.4). Let each security barrier be represented by a pair of nodes connected by an edge. The weight associated with each edge characterizes the time to breach that barrier. Further, we will assume that there is no detection system in place so only the minimum time for an intruder to progress from start to finish will be considered. Once this foundation is in place, the analysis is extended to include detection and response force reaction time.

Stochastic Shortest Path

The first variation to be discussed assumes that we know the distribution of edge weights. The second variation will relax this assumption and simply assume some initial prior distribution on the weights; these edge weights will then be updated as new information becomes available.

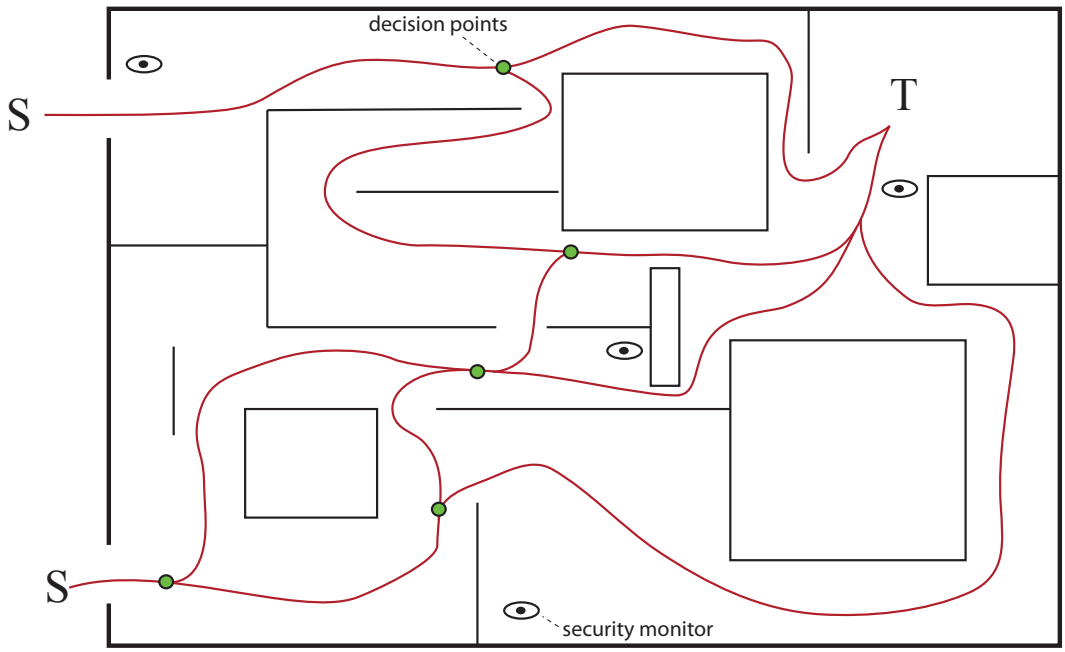

Figure 2.4. Multiple Attack Paths

A graph G is defined by a set of nodes S and a set of directed edges $E \subseteq S \times S$. Every edge $(i, j) \in E$ has a weight of m_{ij}. These edge weights capture the time and difficulty associated with traveling from point i and point j. The graph is assumed to be acyclic, with each edge appearing only once in a given path.

Let $L_p = \sum m_{ij}$ be the length of path p in G that connects the starting node s and termination node t . In this case, the edge weights m_{ij} of G will eventually be considered random variables, where L_p is a random variable with mean and variance, respectively, μ_p and σ_p^2.

Our goal is to find the path p with the smallest total value

$$\min_p \sum_{(i,j)\in E} \delta_{ij}^p m_{ij}$$

where δ_{ij}^p is an indicator function with value 1 if the edge (i, j) is part of the path p and 0 otherwise.

Deterministic Edge Weights

A classic algorithm for identifying all the shortest paths in a network with known edge weights is Dijkstra's approach presented in Appendix A, Algorithm A.1.1 . However, since we are interested only in the shortest path from node s to node t, we can start with s and terminate the algorithm when node t is visited.

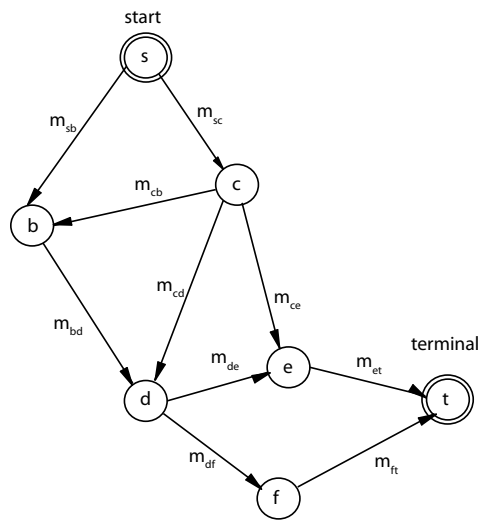

Figure 2.5. Typical Directed Acyclic Graph

Currently, to characterize the security of a particular site, the intruder capabilities are assumed to be known. For example, exercises are conducted and the times to breach a barrier are collected. The set of *best case* times, e.g., 90-th percentiles, are used to populate the graph. A shortest path algorithm, such as Dijkstra's, is used to assess the time to breach. No intercept force reaction times are included in the analysis.

Random Edge Weights with Known Distributions

This first case will follow the stochastic path analysis developed by Deng and Wong [15]. We will assume that m_{ij} are random variables and that the distributions are known or can be estimated. Given the mean and variance of L_p, μ_p and σ_p^2 respectively, it is desired to minimize the cost function $\mu_p + \Phi(\sigma_p^2)$, where $\Phi(\sigma_p^2)$ is an arbitrary function. To assure that the variance in our result is bounded, we will assume an upper bound on the path variance: $\sigma_p^2 < B$. One final minor restriction on the variance is to require the individual variances to be integer values. (It will generally be possible to discretize the variances to the point where this is not a significant restriction.)

Let X_e be a random variable associated with the edge weights. If we assume that the edge weights are all independent with mean and variance μ_e and σ_e^2 respectively, then: $\mu_p = \sum_{e \in P} \mu_e$ and $\sigma_p^2 = \sum_{e \in P} \sigma_e^2$.

If the cost function is of the form $\Phi(\sigma_p^2) = \mu_p + k\sigma_p^2$ then the optimal solution can be found solving the deterministic shortest path problem with edge weights $\mu_e + k\sigma_e^2$. However, the cost function $\mu_p + k\sigma_p^2$ is difficult to relate to the effectiveness of the optimal solution. For example, assume that the intruder breach times are normally distributed random variables and we wish to

characterize the 1% risk of an intruder successfully using the fastest breach time then: $P(|L_p - \mu_p| \leq 3\sigma_p) > 0.99$. This is not possible since the standard deviations are not additive: $\sigma_p \neq \sum_{e \in P} \sigma_e$. It is possible to generalize this beyond Gaussian edge weights by using Chebyshev's Inequality:

$$P(|L_p - \mu_p| \leq k\sigma_p) > 1 - \frac{1}{k^2}.$$

The critical result is that the optimal path is not necessary composed of optimal subpaths. The traditional shortest path algorithms, such as Dijkstra's Algorithm, are no longer applicable.

In this report it will be assumed that the the shortest path is already identified, either through expert judgement or application of the previous approach. However, an unstated requirement for any new approach is that it be possible to extend and account for multiple paths.

2.4.2 Monte Carlo Discrete Event Simulation

Another method that has been used to evaluate security systems is to use Monte Carlo Discrete Event Simulation to track adversaries through a series of random walks through the model of the security system. One can use arbitrary statistical distributions to represent any of the times or probabilities in the model, and include arbitrary conditionality within the model to represent any dependencies that exist in the real system. The advantage of using a method like this is that it can, at least in theory, model the behavior of the system during attack with complete fidelity to the actual system. There have even been simulation tools that model not only P_I, but also attempt to model the actual combat that would occur when neutralizing the adversary force after interruption occurs [21, 2, 3]. While there is considerable debate about the validity of the combat model results, some confidence exists for the path discovery that occurs and the associated computation of P_I.

The chief advantage of this method is the degree to which it can be customized to model the behavior of the system with fidelity that may not be available for the models based on ASDs and directed graphs. This includes the ability do use arbitrary statistical distributions to most appropriately represent the detection and delay data relevant to the system. The disadvantages of the method revolve around its Monte Carlo nature. It is by nature computationally intensive, and because it discovers paths by a random walk, it is impossible to state with certainty that the most attractive attack paths have been discovered and are well-characterized in any particular execution of the software. Conversely, the strength of the directed graph and ASD approaches is their completeness, while they are limited in the fidelity to which they can model the system. It is not unusual for analysts to use these methods in a complementary manner by performing a triage of the system using a tool like ASSESS or ATLAS to develop candidate attractive attack pathways for further analysis using a higher fidelity tool.

2.5 Need for a New Approach

The traditional approaches for evaluating physical security systems possess a variety of advantages and disadvantages as outlined in the preceding sections. However, each of these methods begins with the assumption that the underlying task duration and detection probability data is well-represented using either point values or statistical distributions. In fact, the data are characterized by few experimental trials, sparse observations, and experiments with only partial applicability to the specific conditions under consideration in the facility being analyzed. Indeed, for some of the experiments with multiple trials, the data suffer from a human learning phenomenon that biases later experimental trials to have shorter durations than earlier trials. This is because, while an adversary typically only has one chance to perform a task on the real system, the surrogate adversary in the experiment may have the opportunity to perform the task multiple times for successive experimental trials, effectively learning over the course of the trials and improving both the likelihood of success and the duration of the task. While an adversary may also be able to rehearse on a mockup of the facility to be attacked, there may be considerable uncertainty as to the fidelity with which the mockup represents the real system due to limitations in adversary intelligence gathering.

The limitations in the supporting data are key to the need for a new approach to estimating P_I for security systems and attack pathways. The original experiments and statistical data analyses were performed more than 30 years ago, and modern statistical methods, especially Bayesian statistics and Markov Chain Monte Carlo computational methods, have made it possible to perform computations that are less limited by the conservatisms inherent in previous analysis methods. This project has developed new methods that make optimal use of the existing data, explicitly enable the proper use of data that has only partial applicability, and appropriately captures the uncertainty in the analysis results for parameters such as P_I. The remainder of this report describes this method and provides supporting computational examples. It also provides important insights into how human performance can vary depending on the specific activities and environment of the task being performed, and describes how these factors can be accounted for in the security analysis.

Chapter 3

Introduction to Bayesian Methods

3.1 Background

Regardless of the particular perspective on system security, the critical point is the recognition that resources expended in each level of security should be driven by the impact on the overall level of facility security. The overriding goal is to assess the impact of various uncertainties associated with internal and external variables on the uncertainty in security: an estimate in the probability that the security system will perform as required (P_E, or its component elements P_I, and P_N), in conjunction with the confidence that we have in that estimate.

Clearly a balance must be obtained between the information gained from testing and modeling at various security levels, the resources required to perform testing and develop models, and the uncertainties associated with predicting system performance. The ability to logically combine information from these various areas as well as the organization, characterization and quantification of this myriad of uncertainties are critical elements of the security design and analysis processes. Importantly, although the ultimate objective is to obtain accurate predictions, it must be recognized that merit also exists for simply obtaining more robust solutions than are presently possible (i.e. simply improving the confidence in the facility security estimates).

3.2 An Alternative Approach

Fundamental to implementing the new approach is the recognized need to consider all the evidence at hand through the use of nondeterministic methods. Classical statistical methods, those based on a frequency approach, permit the combination of point estimates but not confidence intervals. Alternatively, the use of a Bayesian methodology provides a formal process for synthesizing and learning from the data that is in agreement with the current information sciences. It permits relevant expert opinion, sparse field data, and laboratory test information to be objectively codified and integrated.

In addition to providing an objective means of combining information, Bayesian methods provide a structured tool for robust model development. For example, if laboratory experiments and detailed mathematical simulations are too expensive, empirical models based on data from field

measurements and/or component/subsystem testing can drive model refinement.

The foundation for what is referred to today as Bayesian analysis was first formalized by Presbyterian minister Thomas Bayes of Tunbridge Wells, Kent. His article Essay Towards Solving a Problem in the Doctrine of Chances was published posthumously in 1763 in the Philosophical Transactions of the Royal Society of London. However, the work was rediscovered in 1774 by Pierre Laplace and it is his formulation that is commonly referred to today as Bayes Theorem.

In the early 20th century, mathematicians became uncomfortable with the concept of only using available information to characterize uncertainty. The concept of long-run relative frequency of an event based on an infinite number of trials became popularized by researchers such as Kolomogorov, Fischer, and Neyman, among others. Practitioners of this latter mathematical philosophy are commonly referred to as Frequentists. However, since it was actually impossible to observe an infinite number of trials, various data transformations (i.e. statistics) were developed. The result has been a array of different statistics for different applications, leading to the classical type of statistics commonly practiced today.

At about this same time, Jefferys[19] published one of the first modern books on Bayes theory, uncovering the work by Laplace many years earlier. Unfortunately, the work appeared at the same time maximum likelihood estimation was being popularized in the literature and much of the potential impact was not realized.

In recent years, the work of Bayes, Laplace and Jefferys has received increasing attention for a wide variety of related reasons. Information sciences and the rise of decision theory as an aid to management are the most recent additions to the list. For engineers, the high cost of testing has led to a push to investigate some alternative means to extort as much information as possible from existing data.

Beyond costly system level testing, possible information sources include materials testing, subsystem laboratory testing as well as data from similar systems and finally computer simulations. From the perspective of increasing confidence in the final decision (within budget constraints), it is of immense importance to efficiently use information from all available sources. For example, information from field analysis and/or laboratory testing at the task level is commonly used to make system level assessments, since laboratory testing is generally much less expensive than full facility testing in an operational environment. This approach has been particularly appealing for those situations where extensive system testing is impractical for a variety of reasons (e.g. cost of prototype development, safety or environmental contraints) or even impossible due to international treaties (e.g. nuclear weapons). While there is no substitute for full scale testing in a realistic operational environment, it is always difficult to justify not considering data from all relevant sources.

In Bayesian statistics, probability is interpreted as representing the degree of belief in a proposition, such as the mean of X is 0.44, or the polar ice cap will melt in 2020, or the polar ice cap would have melted in 2000 if we had not..., etc. Thus we see it can be applied to reasoning about one time events (ice cap melting), counterfactual events (ice cap would have melted), as well as more traditional statistical questions, such as computing distributions over random variables. Bayes rule provides the mechanism by which prior beliefs are converted into posterior beliefs when new data

arrives.

In frequentist statistics (also called classical statistics or orthodox statistics), probability is interpreted as representing long run frequencies of repeatable events. Thus it cannot be used to reason about one time events or counterfactual events. One can talk about the probability of data having a certain value, $p(D|\theta)$ (this is the likelihood function), since one can imagine repeating the experiment and observing different data. But one cannot talk about the probability of a parameter having a certain value, $p(\theta|D)$, since parameters are assumed to be fixed (but unknown) constants, which do not have probability distributions associated with them.

The Bayesian approach is often criticized because the interpretation of probability in terms of beliefs seems subjective. In particular, the dependence on the prior distribution (see below) is seen as unscientific. However, all statistical modeling depends on prior assumptions (e.g., the form of the model); Bayesian statistics make such assumptions explicit.

3.3 Basic Concepts

The subsequent discussion closely follows Wikle and Berliner [33]. Let X be unobservable quantities of interest and D be our data. Then the full probability model can be factored: $P(x,d) = P(x|d)P(d) = P(d|x)P(x)$. Applying Bayes' Rule:

$$P(x|d) = \frac{P(d|x)P(x)}{P(d)} \tag{3.1}$$

$P(d|x)$ is the sampling distribution also known in classical statistics as the *likelihood function*. It represents the data (observations) conditioned on the true, but unobservable, variables. If X is the *true* air temperature, then D represents our observed (possibly inaccurate) measurements of the temperature.

$P(x)$ is the prior distribution. This distribution quantifies our *a priori* understanding of the unobservable quantities of interest. This prior distribution may be based on historical information (climatology) or perhaps from a forecast model. In general, prior distributions can be informative or non-informative, *subjective* or *objective*. The choice of such distributions is an integral part of Bayesian inference.

$P(d)$ is the marginal likelihood distribution where, for the continuous case: $p(d) = \int p(d|x)p(x)dx$. This is simply a normalizing constant that assures that the final distribution is a valid probability distribution.

$P(x|d)$ is the posterior distribution and characterizes our unobservable variables given our data. It is the analysis item of primary interest. It is proportional to the product of the data model and the prior. The posterior is the update of our prior knowledge about X as summarized in $p(x)$ given the actual observations d. In this sense, the Bayesian approach is inherently

scientific in that it is analogous to the scientific method: one has prior belief (information), collects data, and then updates that belief given the new data (information).

We can see from Equation 3.1 that Bayesian analysis is identical to traditional statistical analysis except that the likelihood function, $P(d|x)$, is weighted with the available prior information $P(d)$ (depicted in Figure 3.1).

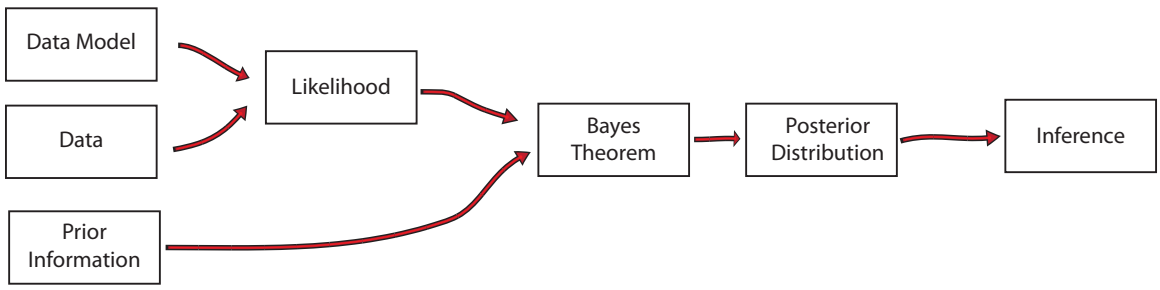

Figure 3.1. Flow of Information in Bayesian Analysis

A major historical impediment to the use of Bayesian techniques is the specification of the appropriate form of the prior. It should be noted that the choice of likelihood function in classical statistics is in many ways also subjective.

As outlined in Berger [7] and Bernardo [8], there are a variety of techniques for developing priors in practice. Priors can be obtained from past studies, from subject area expert opinion or scientific first principles, though compromises are typically made for mathematical simplicity. For example, conjugate prior and data model pairs of distributions lead to posterior distributions of the same functional form as the prior, as with the normal data, normal prior examples above. In other words, in such cases, one knows the normalizing constant in Bayes rule.

There is also a significant literature on the notion of non-informative or objective priors. A non-informative prior does not represent total ignorance about the problem at hand, but is used when information is lacking or very vague. There are typically a variety of non-informative priors in a given problem, which serve as reference or default priors. They also often yield results which match or nearly match those of traditional non- Bayesian approaches, thereby providing them with Bayesian interpretations.

The following section provides a simple example to explore the impact of combining test data with the prior information using a Bayesian updating.

3.4 Example

Appendix B provides a basic introduction to the mathematics underlying Bayes' Theorem with some basic but familiar problems. This section goes a step further and demonstrates the impact of new data on overwhelming the prior assumptions.

Consider Figure 3.2. Assume we are asked to characterize the time required to perform a breach task. Based on historical data (observations of similar tasks, expert judgement, etc.), we can assign a prior distribution on the task completion time with a mean of $\mu_{prior} = 30$ and a standard deviation of $\sigma_{prior} = 8$ (dashed blue line).

We then perform 10 new tests which provide a better estimate of the task time with a mean of $\mu_{data} = 15$ and standard deviation of $\sigma_{data} = 7$ (black line). If we assume that these times are normal distributed random variables then it is rather straightforward to combine the new test data with the historical information using Bayes theorem.

The resulting posterior mean and standard deviation are $\mu_{post} = 16.1$ and $\sigma_{post} = 5.3$ (green line). We can see that the posterior estimates are heavily influenced by the data and overwhelm our initial prior data. However, if we reduce the number of test samples to $N = 1$, and from the data estimate a new mean and standard deviation: $\mu_{data} = 15$, and $\sigma_{data} = 10$, then the prior maintains a stronger influence on the results (red line): $\mu_{post} = 24.1$ and $\sigma_{post} = 6.2$. Note how σ_{post} decreases when the larger number of tests also decreases σ_{data} as also suggested by classical statistics.

3.5 Risk and Benefits of the New Approach

3.5.1 Risk: Establishing Priors

A major point of extraneous contention regarding the use of Bayesian methods is construction of the various priors. It is true that the choice of a poor prior distribution can lead to misleading results. However, the same pitfall occurs with traditional methods where the poor choice of any probability model can lead to poor results. A benefit of Bayesian methods is that they *provide results that are more robust to the choice of the underlying probability model than if traditional approaches were used in isolation.*

3.5.2 Benefits

In addition to the additional level of robustness, there are other benefits to the proposed approach; the full discussion and demonstration of these is beyond the scope of this work. However, a few of the benefits include:

- Objectively deals with limited test data

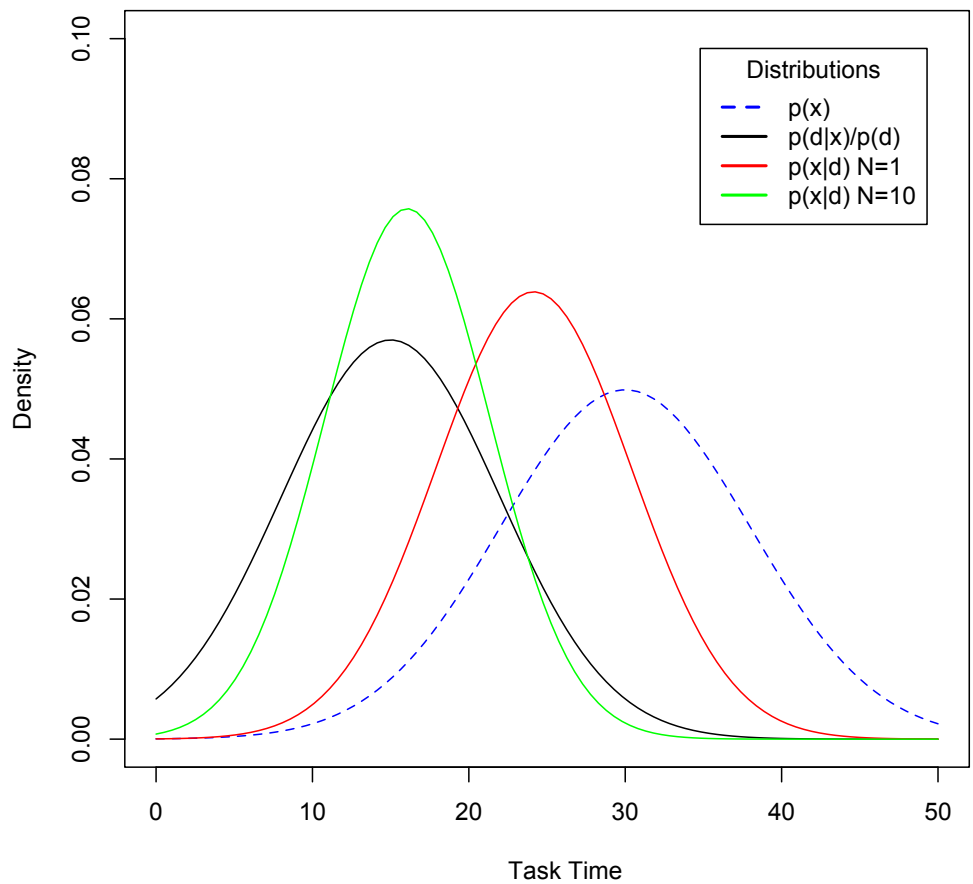

Figure 3.2. Impact of Combining Test Data with Prior

- More cost effective test planning. Data collection is driven by the value of the information gained from the test(s). The new approach can provide direction relative to specific tests and how many samples.

- Objectively consider uncertainties in intruder capabilities. A major uncertainty in facility security assessment is the skill level, risk attitude of the intruder, and possible equipment capabilities. These can be handled directly through the prior distributions and updated as new intelligence or test results become available (Figure 3.3).

- Expert judgement regarding situations that are difficult or costly to test can be incorporated in the analysis.

- Objectively merge data from multiple, possibly similar tests. If we have a series of tests performed by different agencies, under similar but different conditions, it is possible to merge test results and increase the confidence in our security assessment(see the brief discussion of integrating test results in Appendix B.1).

- Deal with unusual situations such as non-zero minimum breach time by:
  - adjusted distribution, e.g. Weibull with non-zero minimum time
  - truncated distribution e.g. truncated normal
  - sum of two distributions, e.g. $T_{min} \sim U(a,b) + T_{task} \sim Weibull(\nu, \eta)$

- Robust to black swan events. No method can predicts or account for extremely rare events. Traditional analysis methods are limited to addressing only those events that can be tested and/or observed. However, if objectively applied, Bayesian methods will assist in identifying the possibility of these events and provide a solution that is more robust to these events than traditional methods (see also the discussion in the Appendix B.2.1).

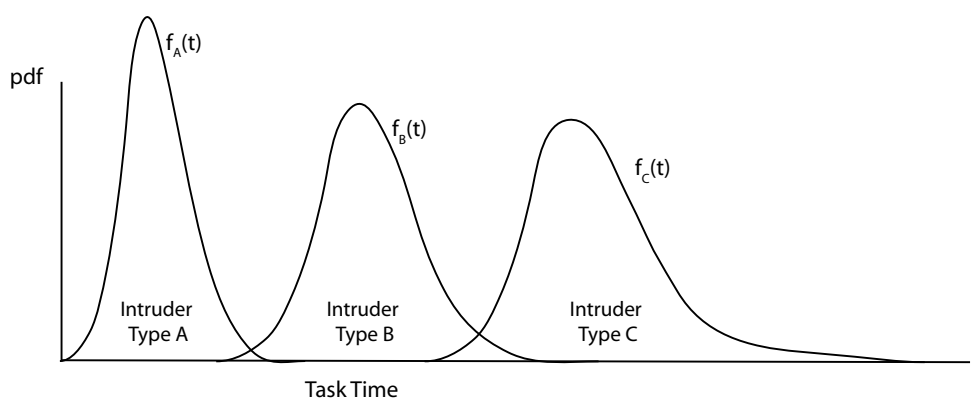

Figure 3.3. Sample Priors Based on Intruder Type

Chapter 4

A New Approach

This section focuses on the development of the basic framework for a new approach. Three very general types of intruder scenarios are introduced and the associated analysis method is developed. The discussion begins with the most basic situation where the intruders are, with certainty, detected when they first breach the facility. The second scenario involves the possibility that during the course of intrusion, various sensors are in place that are each susceptible to failure: detection at each point is not guaranteed. Finally, building on the first two scenarios, an analysis approach is outlined for the situation where the breach of each barrier may involve multiple tasks, each of which must be accomplished before the barrier is successfully breached. Detailed numerical examples of each are provided in the Appendices.

4.1 Introduction

The following discussion assumes that the most likely intruder path has been identified. Either we assume that we already know which path into a facility that an intruder will take or assume that the analysis in the previous section provides a characterization of the stochastic shortest path that an intruder would take through a system. Even in a situation involving multiple likely paths, the underlying approach outlined in the following is extremely fast and could be applied quickly and easily on multiple scenarios.

Figure 4.1 depicts the situation where an intruder moves from barrier to barrier with the possibility of detection at each barrier. Define the following variables:

- n is the length of the shortest path through the security facility, i.e. the maximum number of barriers that must be breach for successful intrusion.
- P_{D_i} probability of detection at point i
- $t_i, i = 1, \ldots, n$ is a random variable describing the time it takes for an intruder to breach barrier i.
- T_p is a random variable characterizing the time it takes for an intruder to successfully traverse all security barriers in the facility, i.e. get from s to t. $T_p = \sum_{i=1}^{n} t_i$.

- t_R is a random variable characterizing the probability that the response force responds by time t_R, cdf $F_R(t)$

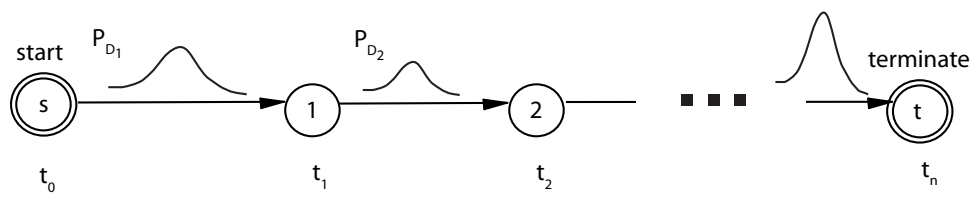

Figure 4.1. Stochastic Shortest Path w/Random Edge Weights

The clock doesn't start for the response team until the intruder is detected. Any intruder detection after the first provides no value (for this single path analysis).

The following sections discuss the results for three different scenarios of increasing complexity and intruder attack resolution:

Case I In this case the intruder is detected at the first barrier and the response team reacts immediately. The analysis then involves, effectively, a race between the intruders and the response team.

Case II The intruder encounters barriers and detection sensors sequentially along a single path. The time to breach each barrier is characterized by a single random variable. The probability of detection at each barrier is a random variable that captures the possibility that a number of barriers may be defeated before the response team becomes aware of their presence. And the race between the intruders and response team begins as in Case I.

Case III This is an extension of Case II with increased resolution of the steps needed by the attack team to breach each of the barriers. Defeat of each barrier requires the successful completion of a sequence of tasks, each of which may require a specific skill set or set of tools.

4.2 Case I: Intruder Detected Immediately

This is the simplest characterization of the problem, removing detection probabilities at each barrier. This simplifies Figure 4.1 to the system depicted in Figure D.2. Assuming intruder detected at first breach, then the response team successfully intercepts the intruders if the response time is less than the passage time for the intruders to breach every barrier: $t_R < T_p$, and the probability of successful site security system performance p_s is given by:

$$P_I = Pr\{\text{response time} < \text{total breach time}\}$$
$$= Pr\left\{t_R < \sum_{k=1}^{n} t_k\right\} = Pr\{t_R < T_p\}$$

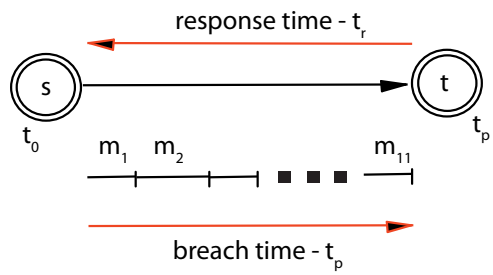

Figure 4.2. Case I: Basic Description

This type of problem is refereed to as an *interference problem* involving the convolution of two distribution functions (Figure 4.3).

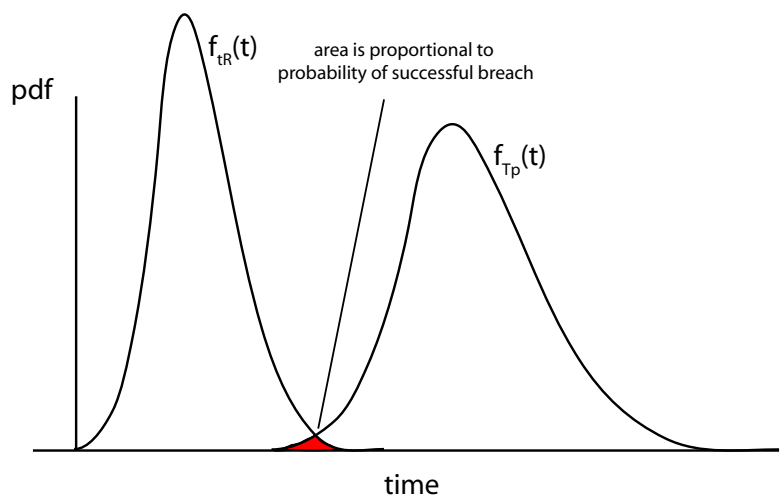

Figure 4.3. Case 1 Distribution Interference

4.3 Case II: Individual Detection Events

Stage 1 Probability that the response team encounters the intruders before they breach all barriers, Pr{response time is less than intruder passage time}. Assuming intruder detected at

first breach, then the response team successfully intercepts the intruders if $t_{R_1} < T_p$, and the probability of successful site security configuration is given by:

$$\begin{aligned} P_{I_1} &= Pr\{\text{detection} \cap \text{fast response}\} \\ &= Pr\{\text{detection}\} Pr\{\text{fast response} | \text{detection}\} \\ P_{I_1} &= P_{D_1} Pr\{t_{R_1} < T_p\} \end{aligned}$$

Stage 2 Intruder not detected at first barrier but detected at second barrier

$$P_{I_2} = [(1 - P_{D_1}) P_{D_2}] Pr\{t_{R_2} < (T_p - t_1)\}$$

Stage 3 Not detected until third barrier

$$P_{I_3} = [(1 - P_{D_1})(1 - P_{D_2}) P_{D_3}] Pr\left\{t_{R_3} < (T_p - (t_1 + t_2))\right\}$$

Stage j Not detected until j^{th} barrier

$$P_{I_j} = P_{D_j} \prod_{i=1}^{j-1} (1 - P_{D_i}) Pr\left\{t_{R_j} < \left[T_p - \sum_{k=1}^{j-1} t_k\right]\right\}$$

$$P_{I_j} = P_{D_j} \prod_{i=1}^{j-1} (1 - P_{D_i}) Pr\left\{t_{R_j} < \sum_{k=j}^{n} t_k\right\}$$

Stage n Not detected until n^{th} (last) barrier

$$P_{I_n} = P_{D_n} \prod_{i=1}^{n} (1 - P_{D_{i-1}}) Pr\left\{t_{R_n} < t_n\right\}\}$$

Finally, the overall probability of adversary success is:

$$P_{AS} = \sum_{j=1}^{n} P_{S_j}$$

because the adversary success scenarios are mutually exclusive. The overall probability of a successful security system:

$$Q_I = \sum_{j=1}^{n} Q_{I_j} = \sum_{j=1}^{n} (1 - P_{I_j})$$

and

$$P_I = 1 - Q_I = 1 - \sum_{j=1}^{n} Q_{I_j} = 1 - \sum_{j=1}^{n} (1 - P_{I_j})$$

4.4 Case III: Intrusion Failure Due to Task Failure

To this point the analysis method has not considered the possibility that the adversary might experience an unrecoverable failure while performing one of the tasks during the attempted attack. That is, the attempted attack has been assumed successful if the adversary is not interrupted before the objective is achieved, regardless of the relative ease or difficulty of the tasks to be accomplished during the attack. For some tasks, the adversary may be able to make repeated attempts if the first attempt is unsuccessful (e.g., tossing a backpack over a fence), with the repeated attempts incurring additional time. Such repeated attempts can be represented in the adversary task time distributions described earlier in this chapter. However, there is also the potential for an adversary to fail at a task in such a manner that the entire attack must be declared a failure (e.g., accidentally stripping delicate bolt threads, power tools that fail to start, or inadvertently causing building collapse when explosively breaching a wall). If the adversary's task list is entirely straightforward, there is little conservatism in assuming that all adversary tasks are performed successfully. But if the task list contains tasks that are delicate and where a second attempt is not possible (i.e., task failure is unrecoverable), the conservatism of this assumption may be large, and a more detailed and realistic calculation may be appropriate.

The previous section described how one can compute the probability of successful site security system performance P_I to interrupt an adversary during an attack timeline that consists of detection events, timed adversary tasks, and a timed security response. The inverse condition, adversary success, was modeled as $1 - P_I = Q_I$. In this extended formulation, the adversary is successful when the following conditions both occur:

- The response force is not timely (i.e., the response force does not successfully intercept the intruders, as described previously), *and*
- The adversary team does not commit any unrecoverable errors.

It is possible to represent this situation using only the method described for Case II if one were to set the time required to perform a task to infinity (or, more practically, to a very large number) to indicate that the adversary could not succeed in performing the task no matter how many times it was attempted. While this may be methodologically simple and intuitively appealing, the probability distributions that would be required to model these task durations would likely be poorly behaved, bimodal, or overly constrained since one would need to ensure that the probability embodied in the tail of the task time distribution (i.e., for those times beyond the response force arrival time) matches the likelihood of the adversary making an unrecoverable error while performing the task. Fortunately, this situation can be modeled simply with an extension to the Case II methodology.

Assume that P_{U_i} represents the probability that the adversary team commits an unrecoverable error while performing the ith task in the attack. Its complement, Q_{U_i}, is the probability that no unrecoverable error is committed during the performance of this task. If the likelihoods for successful completion of these tasks are assumed statistically independent, one can write the an

equation for Q_U, the probability that no unrecoverable errors were committed during any of the n tasks during the attack, as:

$$Q_U = \prod_{i=1}^{n} Q_{U_i} = \prod_{i=1}^{n} (1 - P_{U_i})$$

and P_U the probability that at least one unrecoverable error is committed during the attack, is its complement.

To be successful, the adversary must not make any unrecoverable mistakes Q_{U_i} and must successfully beat the security system (Q_I from the preceding section). Thus, the probability of adversary success for the attack mission P_{AS} can be written as $P_{AS} = Q_I Q_U$, where Q_I is as defined in the preceding section.

Chapter 6 provides demonstration problems in which adversary task times are represented as Bayesian random variables (Case I). Appendix D provides an additional demonstration problem in which both the adversary task times and the detection probabilities are represented as Bayesian random variables (Case II). A calculation in which the likelihoods for unrecoverable errors are also represented as Bayesian random variables (Case III) would follow the same methods as the demonstration problem in Appendix D. Alternatively, one could obtain a scoping estimate for P_{AS} by computing P_I using the Case II equations and initial estimates for P_U to determine whether the additional insights to be gained are worthy of the additional computational complexity required by the Case III equations.

A very simple system is analyzed in Appendix D to demonstrate the methodology.

Chapter 5

Integrating Human Factors

5.1 Background

Several variables may impact human performance on the type of tasks needing to be completed by an adversary during an attack. Fatigue, workload, lighting, noise, extreme temperature, and task switching are some of the stressors that can affect a persons performance within situations similar to those being discussed in this report. However, there is disagreement among researchers regarding the level of disruption such stressors may cause [29]. Furthermore, there is a lack of understanding of what constitutes stress and how to measure it. Dekker [14] explains the issues as, *Psychologists still debate whether stress is a feature of a situation, the mental result of a situation, or a physiological and psychological coping strategy that allows us to deal with a demanding situation. This complicates the use of stress in any causal statement, because what produced what?* (pg. 141). This section presents background on the variables most relevant to the scenarios of interest within this report and further elaborates on the difficulties encountered in trying to model their impacts: workload, ergonomics, fitness for duty, and time pressure.

5.2 Workload

Workload may be defined as either the tasks the environment places on the individual (defined as task load by Parasuraman and Hancock [24]) or as the internal interpretation of the loading by the person in attempting to adapt to the situation. This distinction may appear rather arbitrary however, as the perception, either real or created, will impact the individual in the same way. However, the impact on a persons performance will not be the same across individuals.

Workload often becomes a catch all phrase to refer to multiple forms of stress impacting an individuals performance. Furthermore, the impacts may be positive or negative. Too low a level of workload and an individual will easily become bored. This phenomenon may be particularly relevant during a monitoring task when the person must stay tuned to some instrument or indicator for a long period of time with no other engaging task. After a certain length of time (different across individuals), the person becomes bored and easily distracted or loses all monitoring capability by falling asleep.

The ability to handle the stress of increased (or, conversely, decreased) workload is aided by a persons experience and training. There also exists a level of optimal arousal (or stress) to which a person responds. Figure 5.1 [11, 32, 34] demonstrates the effect upon performance based on level of stress experienced by the individual. Eustress, which represents the facilitative effects of stress, is seen at the optimal arousal level range. Above this optimal level, performance degrades until it reaches a point where performance fails completely due to the effects of distress in the situation. Swain and Guttman [31] make the direct connection between the performance effects of stress to the effect of task or workload. In their description of levels of task load, the effect on performance mirrors that explained by differing effects of stress.

Although in theory the arousal curve could be used to help predict the performance of an adversary during an attack, difficulty is encountered in determining where on the curve an individual would fall given a specific situation. Each individual responds differently in terms of stress level and effect of workload to the situation. Furthermore, the magnitude of the impact is difficult to determine and would also differ across individuals. Therefore, a generalized rule specifying an error adjustment to be applied in the model representing all cases cannot be recommended.

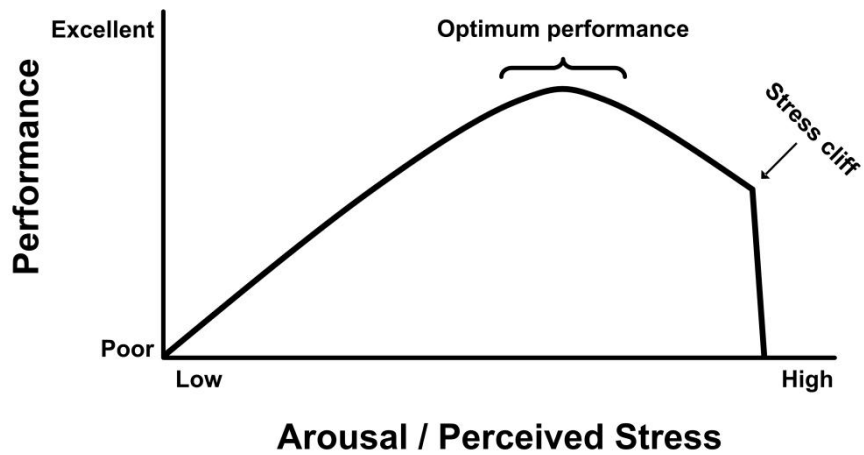

Figure 5.1. Performance versus arousal/perceived stress curve [11]

5.3 Ergonomics

Ergonomic issues could certainly be present during a site breach. The adversary may face a number of obstacles creating an ergonomic challenge such as low lighting (e.g., either through low illumination or high amounts of dust particles in the air following, for example, an explosion), loud noises (especially in the presence of alarms), confined spaces, and/or carrying heavy weight.

Ergonomic issues alone or in combination may serve in two roles that may hamper an adver-

sarys progress. First, these variables create a stressful environment with similar consequences to those already discussed in the workload section. The degraded ergonomic conditions often present a novel situation on which the individual has no experience or training. For instance, although an individual may be quite proficient in operating some tool, once his or her vision is obscured (e.g., through low illumination or thick dust particles in the air), the individual is unable to operate the tool through touch alone.

The nature of the task being executed is defined by the less than optimal conditions. That is, due to the secrecy and desire to go undetected as well as the necessity of working in conditions that were specifically designed to inhibit the task being completed (i.e., breaching an area that was designed to not be breached), the adversary faces ergonomic challenges. As already mentioned, lighting may become an issue either due to time of the attack (many attacks occur during the night), location of attack (sites of breach are in low-light areas to discourage being seen by surveillance), or clouded vision (due to dust particles in the air caused by the breaching activities, or deliberate obscurance is often designed into security systems). Low light (or even no light) creates obvious difficulties for the adversary in completing their tasks. It is only for the advanced attacker (i.e., Level A adversary to be explained in Section 5.7) that lighting would probably not create an obstacle slowing progress as they are typically prepared for such difficulties (e.g., through the use of night-vision goggles).

Working in confined spaces presents another difficulty for the average attacker. This challenge goes beyond the need to often crawl though restricted openings. That particular challenge would be estimated directly in the times to complete the task. However, the variability seen in individuals using a tool or manipulating some item in a confined workspace is not as easily accounted for. Working in such areas may also lead to psychological impairments due to claustrophobic feelings stressing the individual and causing numerous mistakes.

Completing tasks in the presence of loud noises has been shown in many studies to lead to degraded performance [29]. Staal points to several studies showing noise negatively impacting performance in several areas including tasks of vigilance and attention, memory, reaction-time, and psychomotor tasks. Not only might the loud noise interfere with cognitive processing of the task at hand, but if the adversary is working as a team, the noise will interfere with communications between team members. As mentioned previously, these effects by noise are probably more pronounced for the average adversary.

A final ergonomic condition to consider is the common need to carry heavy loads during the attack. These loads may be due to the need to carry tools or, during the exit, carrying the newly acquired prize. The ability to deal with such loads is clearly dependent on the individual. A further complicating factor is the impact on response time and speed the weight has further into the scenario when fatigue is setting in. the effect of fatigue is discussed further in the next section.

5.4 Fitness for Duty

The phrase *fitness for duty* may be used to capture several variables impacting performance. Fatigue and the influence of drugs or medications are two of the most prevalent and evaluated variables, especially regarding performance in a workplace. The impact drugs or medication may have on the individual's performance is highly dependent on the individual as well as the substance. Therefore, this impact is very difficult to predict or model. The effects of fatigue, on the other hand, are well studied and have continuously been shown to negatively impact performance.

Although the effects of fatigue are well described, the mechanism of effect is not so well understood. Cercarelli and Ryan [12] offer the following explanation for fatigue, *fatigue involves a diminished capacity for work and possibly decrements in attention, perception, decision making, and skill performance*. The effects of fatigue may be divided into those that are physical (i.e., muscle tiredness causing mistakes in performance) versus cognitive or mental. It is the mental fatigue that is perhaps more difficult to manage as it is more difficult to recognize and recover from. Mental fatigue may be a factor either due to lack of sleep or due to continuous interaction, monitoring, or cognitive stimulation. Staal [29] points out that one hypothesis attempting to explain the effects of fatigue *proposes that fatigue removes resources in some direct way or perhaps indirectly diverts them toward coping strategies. Thus, performance falls off on tasks due to the depletion of resources* (pg. 93).

Similar to the difficulty experienced in trying to measure stress and predict its effect on performance, fatigue has been a difficult variable to measure. For both phenomena, the difficulty boils down to a difficulty in measuring the variable. Staal [29] captures the issue well as, *In many instances, fatigue is spoken of as being present or absent, yet theoretically most consider it to be a continuous variable. Thus, at what degree of subjective experience or objective measure is one to be considered to be fatigued?* ([29], pg. 94).

5.5 Time Pressure

Time pressure and its effects on the performance of an adversary defeating a barrier are probably the most relevant to this model. Although the other variables may certainly be relevant, the ability to predict their influence is difficult unless an adversay description and some specific factors surrounding the situation are known or can be closely estimated. In fact, many researchers contend that time pressure is at the root of all performance decrements and that any element impacting workload does so through this variable [29],. This observation, and the broad effects of time pressure, point to its importance in being assessed in situations such as those being discussed in this report. Staal's review explains the effects of time pressure have been found to degrade performance in a number of areas such as judgment and decision making, visual search, vigilance and attention, memory recall, and subjects self-rating of performance. The effects of time pressure are far reaching: In addition to a general drop in performance, time pressure and the corresponding sense of urgency experienced tends to result in strategy shifting in teams (explicit to implicit rules and

greater coordination between members), task- or load-shedding (of which strategy-shifting may be seen as one specific example), tunneling of attention and visual scanning, and a speed/accuracy trade-off in performance (pg. 84). Zakey [35] asserts that the degradation seen in performance of complex tasks due to time pressure could be attributed to the shortage of cognitive resources leading to the adoption of simple strategies and an increase in errors. Amabile et al. [5] researched the effects of time pressure on creative thought and found that individuals are less likely to be able to think creatively if under time pressure. This effect may be particularly pertinent to this report as an adversary will likely encounter situations he or she will be required to creatively think through solutions for and the pressure of time will slow his or her ability to do so.

5.6 Accounting for Effect

Human actions are embedded within the contexts in which they occur. It is, therefore, impossible to account for the effect of human error without understanding and accounting for the context within which those errors occur [28]. One manner in which the effects of human performance variables are accounted for in determining failure rates is the study of human reliability analysis (HRA). Within HRA, the context leading to a human failure event (HFE) is described in enough detail to allow the analyst to make a judgment as to the most likely performance influencing factors (PIFs) impacting behavior. These PIFs represent the stressors, similar to the ones described in this chapter, affecting human performance. The impact of the PIFs is accounted for when calculating the human error probability (HEP). PIFs may have a negative or a positive impact on human performance. However, for our purposes, we are more interested in the negative impact and how it might affect the likelihood of success.

Several HRA references may be useful in determining appropriate multipliers to apply to calculations of error rates when defeating barriers (SPAR-H, THERP, etc.). *However, a fairly standard multiplier applied in this arena is two*. Given severe time pressure (e.g., trying to compensate for an error made in an emergency situation), the initial error rate is multiplied by two and is doubled for each additional attempt after the initial failure until the limiting condition of 1.0 is reached (Swain and Guttman [31]. To account for more specific individual-, environmental- or task-related influences on performance, the HRA methods may be used to find specific multipliers or adjustments to apply representing specific PIFs. The HRA method SPAR-H (standardized plant analysis risk-human reliability analysis; Gertman et al. [17]) is based on the assessment of eight key PIFs: available time, stress, complexity, experience and training, quality of procedures, ergonomics and human-machine interface, fitness for duty, and work processes. Boring et al. [10] offers guidance on how to assess the PIFs and the appropriate multipliers to apply. Similar detailed guidance is provided in THERP (Technique for Human Error Rate Prediction; Swain and Guttman [31]) on a few PIFs if information is known on the context of the situation.

An alternative to the use of multipliers in accounting for the impact of human error on performance is to capture within the model the distribution of PIF. Berry [9] argues that the often used Gaussian (or normal) distribution is not the correct distribution for modeling task performance, although it is often used in this capacity. In contract, Berry argues that the Weibull distribution

is more aptly suited for modeling human performance tasks times. However, data is often not available for estimating the Weibull parameters. Berry offers guidance in which the parameters maybe estimated after determining (through expert judgment) the mean and standard deviation of task performance time. This lack of data on human performance in specific situations creates the real problem seen in including PIFs within the model. It is a further complication in estimating performance differences between the three levels of adversaries discussed in the next section.

5.7 Adversary Skill Levels

Three types of adversaries are thought of in relation to the scenarios discussed in this report: types A, B, and C. The Type A adversary is an expert in the field. This individual comes equipped with all the necessary tools, has knowledge on the use of the tools in a variety of situations, and is extremely capable of success in what is considered a stressful environment. Type B adversary has some training in the tasks, but would not be considered a *top of the line* expert. This individual is very capable of success, but he or she would succumb to some stressful impacts. Finally, the Type C adversary is the average individual who would be the most susceptible to the stressful impacts incurred in these tasks. One of the primary difficulties in predicting human performance in an attack breach scenario is understanding what type of adversary is being considered. The optimal case would be to build separate models for each type of adversary. However, this is impractical due to the dearth of data on human performance for any of the adversary types across multiple tasks. Furthermore, there is difficulty in simply defining what category the adversary falls in; that is, given a certain adversary description, what type would he or she be assigned to? Although an adversary may be considered Type A in one task type, does this label apply to the same individual for all tasks involved in the site breach? Finally, it is possible that the type of distribution associated with the performance task time is different across the adversary types, but further research on experts versus novices in these domains is needed.

Although it may be tempting to model the Type A adversary by setting each of the PIFs to a nominal state, thereby taking a very conservative view and stating the adversary is not likely to fail based on human error due to a stressor, this view would be flawed. Although experts excel in a number of ways making their performances more robust than novices, they are susceptible to errors in a number of instances. Expertise is domain-limited such that although an individual may be an expert in one particular area or on some particular task, his or her performance in other areas or tasks may be very similar to those less experienced or novices. Experts may also show inflexibility in *adapting to changes in problems that have a deep structure that deviates from those that are acceptable in the domain* ([13], pg. 26). This inflexibility in an expert's response may also manifest as a bias in response to other similar task types. *Greater domain knowledge can also be deleterious by creating mental set or functional fixedness. In a problem-solving context, there is some suggestion that the more knowledgeable participants exhibit more functional fixedness in that they have more difficult coming up with creative solutions* ([13], pg27). That is, an expert becomes fixated on applying their specialized knowledge to any problem encountered that resembles the one they are most prepared for. This vulnerability may be particularly relevant when predicting performance by experts in these attack scenarios.

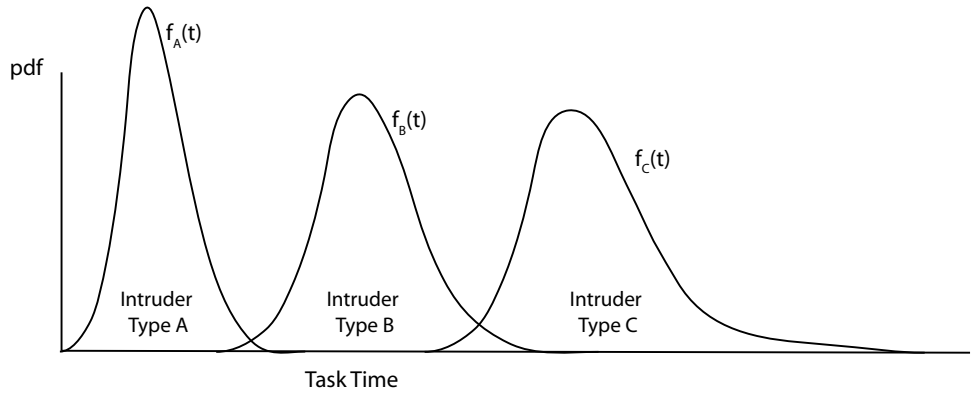

Figure 5.2. Priors Based on Intruder Type

Given the vulnerabilities in an expert's performance, it is inaccurate to think that experts are immune to human performance issues. However, it is likely that the variability in their response times is much more restricted than that for the other two types of adversaries. *The literature suggests that when inter-individual variability declines with the development of expertise, a group of expert performers can be expected to show much smaller differences between them than do novices* ([4], pg. 153). *Therefore, one approach to consider for modeling the performance task time is to let the coefficient of variation be constant across all intruder types.* Then as the mean breach time increases, so does the standard deviation (see Figure 5.2). An alternative perspective would suggest a more general statement: as skill level decreases, the variance in task times increases.

Chapter 6

Demonstration of Approach

This section brings together the major points of the previous material and, using a simple scenario, demonstrates an access delay assessment using the proposed framework. The access delay situation is first analyzed using traditional methods and then the new framework is employed. The two results from the two methods are then compared.

6.1 Problem Description

Consider an access delay situation where there are seven barriers, each with multiple tasks required to breach each barrier. While the data are taken from actual delay testing, the details of what has been tested are withheld for security reasons. Data for this example is provided in Appendix E. In some cases, there were resources available for twelve tests, while for others, only one test was possible. Of the twenty-eight total tasks, the data recorded for each task vary between one and twelve repetitions, with most tasks receiving three to five. In summary, the adversary will be required to complete all 28 tasks in order to complete their objective.

The issues become: how to combine the test data from each task into a characterization of the breach time for that barrier and how to *objectively* combine the breach times for the barriers into a total time to breach the security system for the facility.

6.2 Traditional Access Delay Calculation

For the traditional access delay calculation, two calculations are presented. The first is the calculated delay using the average of the measured data for each task. The total accumulated time for this scenario is 3227 seconds (53.8 minutes) (Table 6.1) .

The second scenario calculates the accumulated delay from 50% of the average measured times for each task. The total time for this scenario is 1614 seconds (26.9 minutes) (Table 6.2).

Table 6.1. Path Time for Average

Barrier	Task	Task Time	Accumulated Time	Comments
1	1	66.7	66.7	Barrier 1, Two Tasks, 173 sec
	2	106.3	173	
2	3	16.2	189.2	Barrier 2, One task, 16.2 sec
3	4	67.8	257	Barrier 3, Five Tasks, 2279.1 sec
	5	718	975	
	6	1051	2026	
	7	313	2339	
	8	129.3	2468.3	
4	9	17.4	2485.7	Barrier 4, Five Tasks, 210.8 sec
	10	26.8	2512.5	
	11	37.2	2549.7	
	12	57.6	2607.3	
	13	71.8	2679.1	
5	14	54	2733.1	Barrier 5, Five Tasks, 420.1 sec
	15	71.6	2804.7	
	16	84.5	2889.2	
	17	90.8	2980	
	18	119.3	3099.2	
6	19	5.7	3104.9	Barrier 6, Three Tasks, 25.7 sec
	20	8	3112.9	
	21	12	3124.9	
7	22	8	3132.9	Barrier 7, Seven Tasks, 102 sec
	23	3.8	3136.7	
	24	9	3145.7	
	25	1	3146.7	
	26	42.5	3189.2	
	27	18	3207.2	
	28	19.8	3226.9	

Table 6.2. Path Time for - 50%

Barrier	Task	Task Time	Accumulated Time	Comments
1	1	33.3	33.3	Barrier 1, Two Tasks, 86.5 sec
	2	53.2	86.5	
2	3	8.1	94.6	Barrier 2, One task, 8.1 sec
3	4	33.9	128.5	Barrier 3, Five Tasks, 1139.6 sec
	5	359	487.5	
	6	525.5	1013	
	7	156.5	1169.5	
	8	64.7	1234.2	
4	9	8.7	1242.9	Barrier 4, Five Tasks, 105.4 sec
	10	13.4	1256.3	
	11	18.6	1274.9	
	12	28.8	1303.7	
	13	35.9	1339.6	
5	14	27	1366.6	Barrier 5, Five Tasks, 210.1 sec
	15	35.8	1402.4	
	16	42.3	1444.6	
	17	45.4	1490	
	18	59.6	1549.6	
6	19	2.9	1552.5	Barrier 6, Three Tasks, 12.9 sec
	20	4	1556.5	
	21	6	1562.5	
7	22	4	1566.5	Barrier 7, Seven Tasks, 51 sec
	23	1.9	1568.3	
	24	4.5	1572.8	
	25	0.5	1573.3	
	26	21.3	1594.6	
	27	9	1603.6	
	28	9.9	1613.5	

6.3 Statistical Analysis

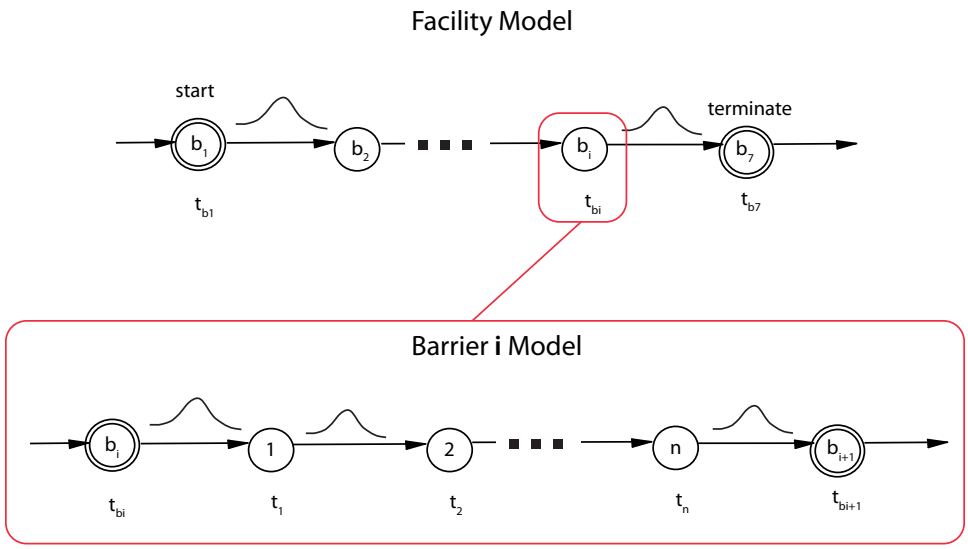

Figure 6.1. Facility Barrier-Task Decomposition

6.3.1 Task Times

For this example, we will assume that the time required to complete each task is a lognormally distributed random variable (Figure 6.1):

$$f(t|\mu,\tau) = \sqrt{\frac{\tau}{2\pi t^2}} \exp\left[-\frac{\tau}{2}(\log(t)-\lambda)^2\right], \qquad t > 0 \tag{6.1}$$

The lognormal is a logical choice since it applies only for task times > 0 and similar to the Weibull distribution, with two parameters, it can take on a variety of shapes. Further, it can easily be modified; for example truncated to allow for situations where a constant or minimum time is required to complete a task.

It can be shown that if $T \sim$lognormal, then the transformed random variable $ln(T) = Z \sim$ normal.

$$\begin{aligned}
E(Z) = \mu_Z &= \lambda \\
V(Z) = \sigma_Z^2 &= 1/\tau \\
E(T) = \mu_T &= \exp[\lambda + 1/2\tau] \\
V(T) = \sigma_T^2 &= \mu_T^2[\exp(1/\tau) - 1]
\end{aligned}$$

Recall that μ_T, σ_T^2 are random variables, and therefore μ_Z, σ_Z^2 are random variables. The next section outlines an approach for estimating the prior distribution for these parameters to be used to initiate our analysis.

6.3.2 Construction of Priors

While our survival is intimately tied to causality and probability theory, humans continue to struggle with communicating and summarizing probability. This can present a challenge when establishing priors for our analysis. Our goal then is to: 1) make the establishing the prior as objective as possible, 2) ensure that our results are insensitive to the prior.

However, we are probably most comfortable with the normal or Gaussian distribution. Since $Z = ln(T) \sim N(\mu_Z, \sigma_Z^2)$, we will form our priors in Z-space and then easily transform the priors back to T-space.

While the following may appear to be rather complicated, in practice it is straightforward to automate within a security analysis program and would require very limited user input.

Consider the test results for Barrier 1 in Appendix E and for convenience repeated below in Table 6.3. Let $\mu_Z \sim N(\mu, \sigma)$. A reasonable estimate of $\mu = E[\mu_Z]$ is the mean of the transformed

Table 6.3. Barrier 1 Data

Test	Task 1	Task 2
1	75	113
2	64	106
3	61	100
μ	66.7	106.3
σ	7.37	6.51

task times from our historical data: $\mu = \sum_i \ln(t_i)/n \approx 4.19$. The choice of $V[\mu_Z] = \sigma^2$ can potentially influence the sensitivity of our results to our prior on the mean; we want this sensitivity to be as low as possible. For this reason will will set σ^2 to a very large (relative) value on the order of $(3\mu)^2 \approx 144$. The result is a prior distribution centered on the mean from the test data, but with a very flat shape across the interval of extreme possible values: $(\mu - 3\mu, \mu + 3\mu) = (-8, 16)$ (Figure 6.3).

To estimate a prior on σ_Z^2 we will assume that the distribution for σ_Z^2 is an inverse-gamma distribution: $\sigma_Z^2 \sim InvGamma(\alpha, \beta)$ where $E[\sigma_Z^2] = \beta/(\alpha - 1)$ and $V[\sigma_Z^2] = \beta^2/[(\alpha-1)^2(\alpha-2)]$. Fortunately, it's easier to find a prior on $\tau = 1/\sigma_Z^2$. If σ_Z^2 follows an inverse gamma distribution, then τ is a gamma distribution with $E[\tau_Z] = \alpha/\beta$ and $V[\tau_Z] = \alpha/\beta^2$. Common robust values for our problem are $\alpha = 1$ and $\beta = 1/(2\mu)$; for Task 1 $\mu = 4.2$.

Therefore the expected variance is $E[\sigma_Z^2] = 0.112$ and variance of the variance is $V[\sigma_Z^2] = 0.014$.

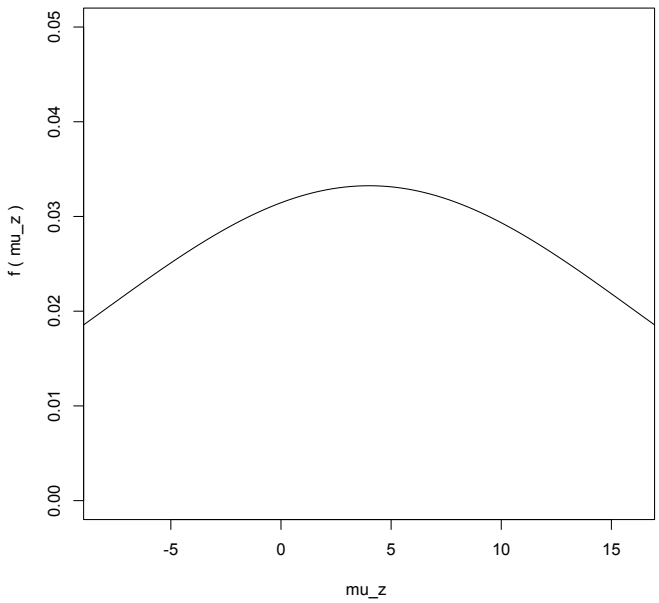

Figure 6.2. Prior on $\mu_z = E[ln(t)]$

An alternative that provides similar results is to relax the previously assumed independence between μ and σ and assign a prior on $\mu|\sigma^2 \sim N(\mu_Z, \sigma^2/n_0)$, and $\sigma^2 \sim InvGamma(\nu_0/2, \nu_0\sigma_0^2/2)$ where $\nu_0 > 0$ is chosen to be the number of historical samples to find the sample variance $\nu_0\sigma_0^2$ from the historical sample, and n_0 is a prior 'sample size' that characterized the strength of belief in the prior.)

6.4 Results

Analysis of the access delay timeline was accomplished using a open source Markov Chain Monte Carlo simulation tool (JAGS [25]).

The following tables summarize the uncertainty analysis of the time to complete each task ($zt[i]$ in Table 6.4) and uncertainty in the time to breach the individual barriers (b_i in Table 6.4). For each variable, the 2.5, 25, 50 (median), 75, and 97.5-percent quantiles are provided. *These quantiles are not the same as confidence intervals.* These quantiles are estimates of the probability that the variable of interest is less than the specified value. For example, in Table 6.4 we see that $P[\text{total delay} < 14311.4 \text{ sec}] = 0.975$.

From Table 6.6 we can see that the median times correspond well with the results from the traditional method of only using the mean values (also evident in Figure 6.5). However, as we can

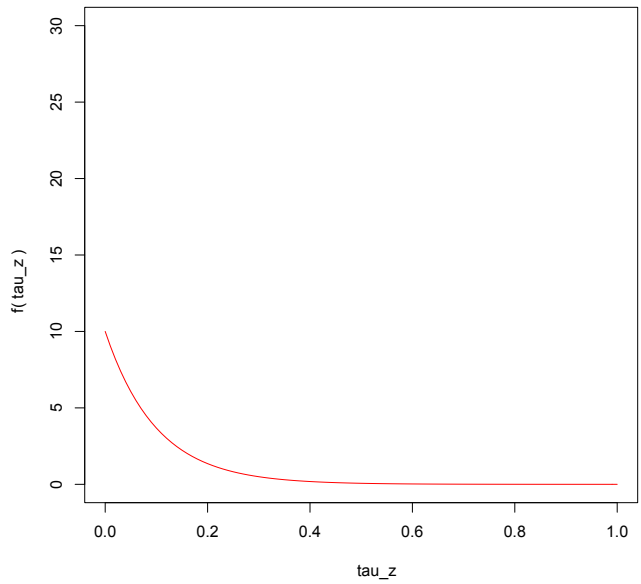

Figure 6.3. Prior on $\tau_z = 1/V[ln(t)]$

see from Table 6.4 there is considerable uncertainty in the total access delay time. Since there is such limited task data available for some of the barriers, and for others the test times for a single task vary by as much as 200%, this result is not surprising.

As a rough comparison, it can be seen that the Bayesian and the traditional methods provide similar results when the mean and the median times to overwhelm all barriers are compared: 3453 versus 3227. However, the range for the new approach and the traditional approach are compared, the differences of the two methods is clear: $[2295, 14311]$ versus $[1614, 4841]$. The new approach provides a broader, hopefully more realistic characterization of the intruder delay.

Figure 6.6 provides an additional perspective. As noted previously, there is uncertainty in the Response Force reaction time. Simultaneously considering the facility access time, T_A, by the Adversary and the Response Force time, T_R, the probability that Response Team intercepts the Adversary, $T_A - T_R > 0$ is depicted in Figure 6.6.

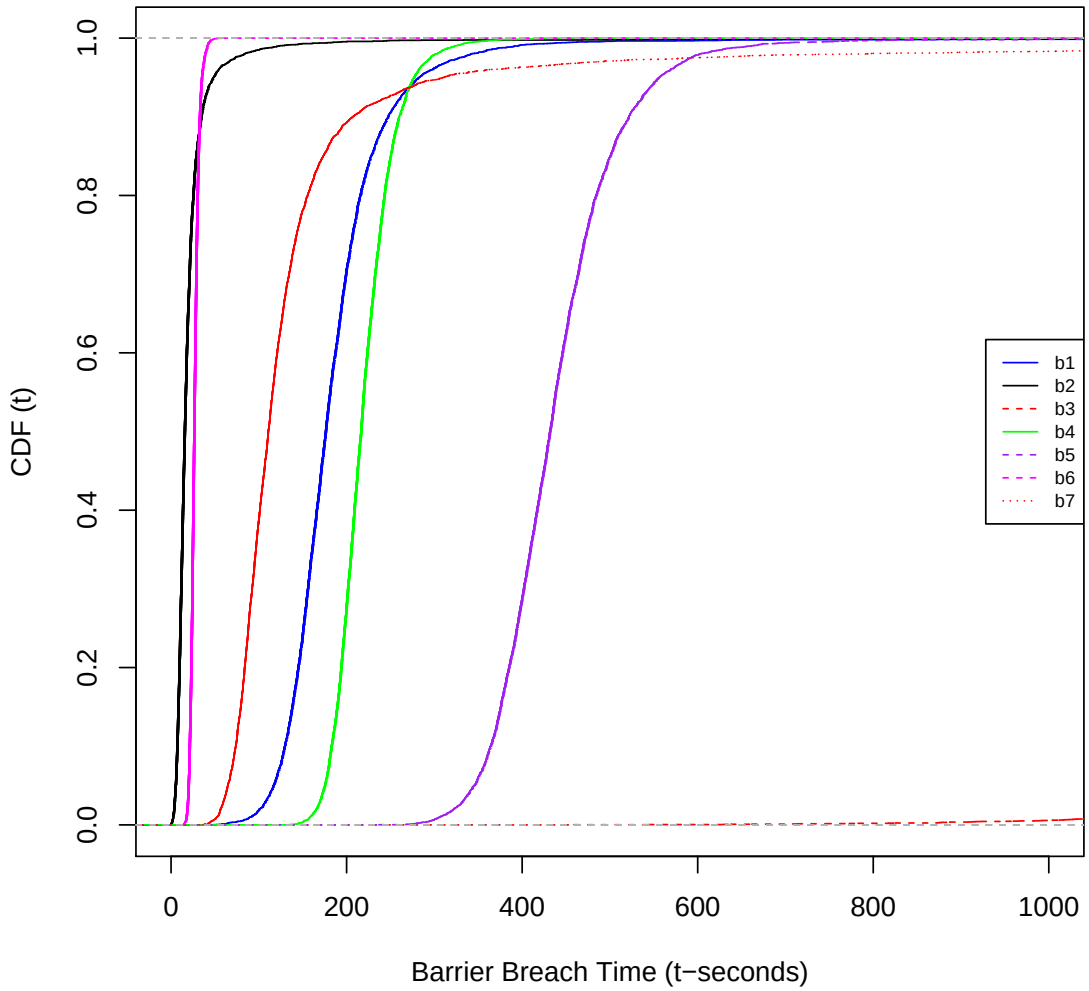

Figure 6.4. CDF for Barrier Breach Times (t-seconds)

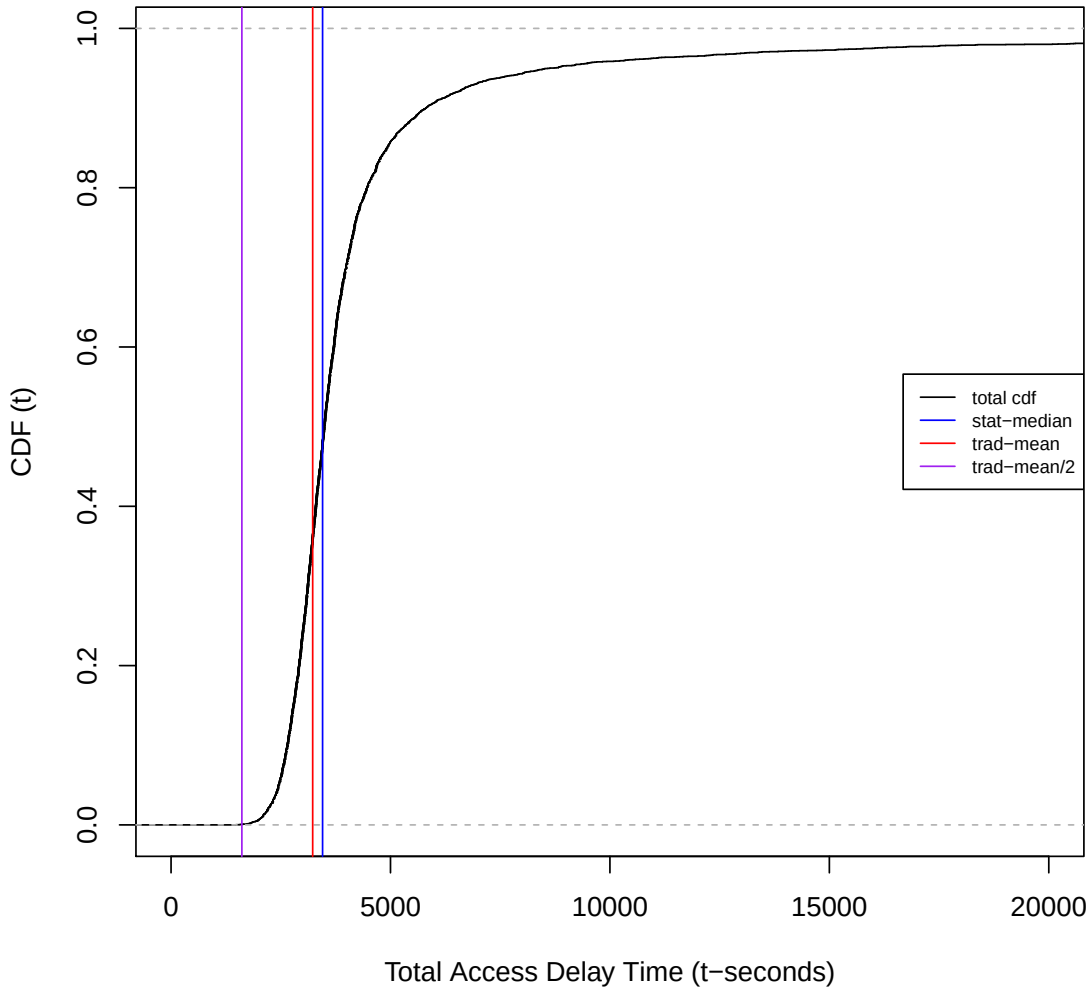

Figure 6.5. CDF - Facility Breach Time (t- seconds)

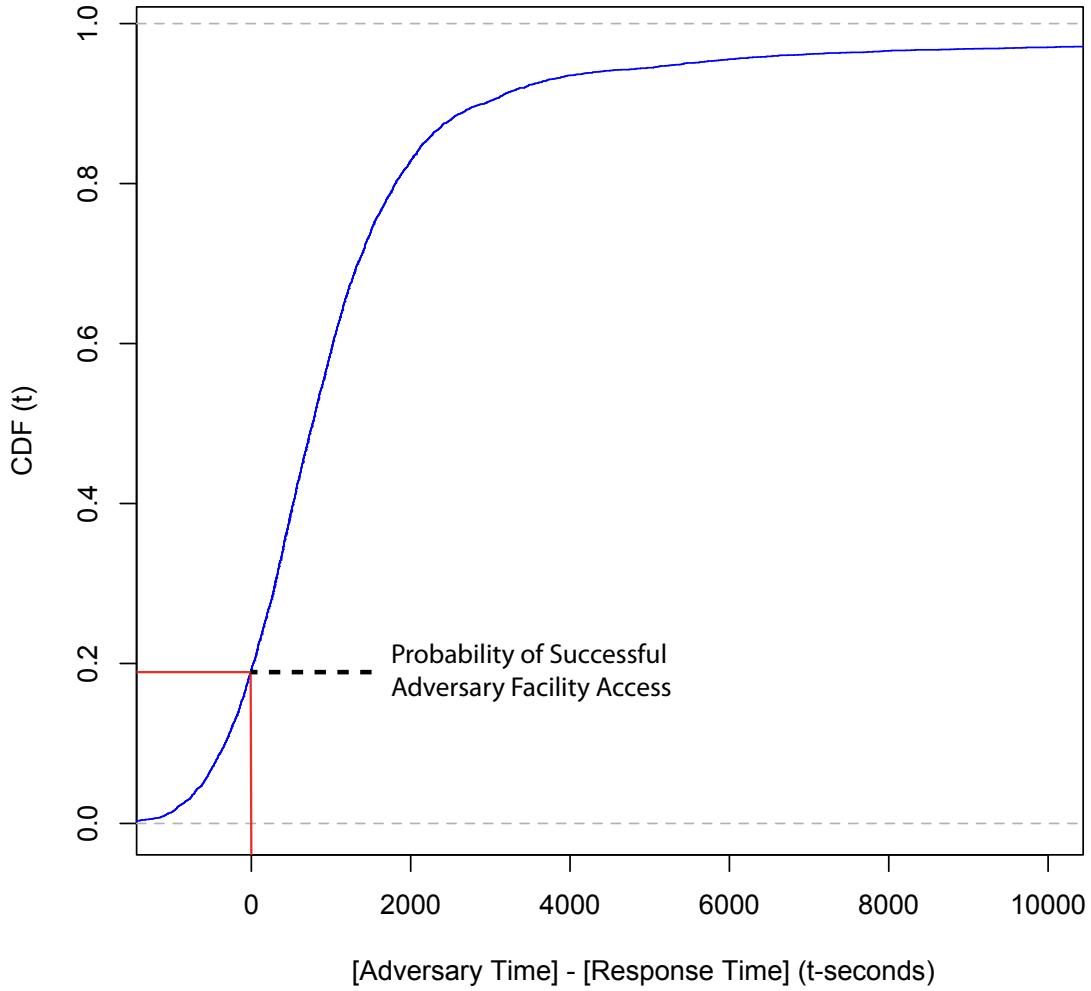

Figure 6.6. Pr{Response Team Intercepts Adversary} (t- seconds)

Table 6.4. Individual Task Times - Quantiles

Task	Quantile 2.5%	25%	50%	75%	97.5%
zt[1]	29.83	53.14	65.99	82.35	146.65
zt[2]	51.11	87.15	105.75	130.25	227.36
zt[3]	3.42	10.08	15.17	22.41	68.72
zt[4]	29.90	53.10	65.88	83.19	140.76
zt[5]	145.09	530.31	722.35	982.29	4177.68
zt[6]	231.50	781.90	1051.27	1439.00	5089.02
zt[7]	117.64	246.46	312.77	393.16	799.23
zt[8]	35.98	87.91	121.77	167.57	420.82
zt[9]	8.62	14.16	17.35	21.26	34.37
zt[10]	14.44	22.24	26.67	32.17	49.81
zt[11]	20.42	31.15	36.96	44.12	67.06
zt[12]	31.89	48.25	57.29	67.97	102.41
zt[13]	41.94	61.55	72.03	84.77	124.36
zt[14]	29.73	45.53	53.71	63.47	95.98
zt[15]	37.79	59.05	70.61	84.78	129.02
zt[16]	43.78	70.42	84.33	101.02	156.07
zt[17]	50.09	75.75	90.56	108.28	165.19
zt[18]	64.17	99.89	118.97	141.63	222.94
zt[19]	2.88	4.51	5.53	6.90	10.95
zt[20]	4.67	6.60	7.91	9.42	13.97
zt[21]	6.80	9.95	11.85	14.14	20.71
zt[22]	3.26	6.22	8.07	10.41	21.42
zt[23]	1.15	2.67	3.74	5.20	12.17
zt[24]	1.57	5.10	8.04	12.48	42.48
zt[25]	0.29	0.71	1.01	1.45	3.62
zt[26]	7.09	25.86	38.76	59.85	245.12
zt[27]	1.42	11.15	18.19	28.88	215.80
zt[28]	8.16	15.01	19.39	24.94	47.28

Table 6.5. Barrier Breach Times - Quantiles

Barrier	2.5%	25%	50%	75%	97.5%
1	104.93	150.77	176.16	207.28	328.66
2	3.42	10.08	15.17	22.41	68.72
3	1302.87	1991.69	2414.29	3079.12	10927.52
4	165.23	197.52	216.12	238.29	294.02
5	329.73	394.36	430.06	473.26	585.42
6	18.51	23.21	25.97	29.11	37.60
7	58.73	88.92	109.87	144.82	581.32
total delay	2295.33	3013.38	3452.57	4165.90	14311.41

Table 6.6. Comparison of Traditional v Statistical

	Traditional		**Statistical**
Barrier	Mean	0.5*Mean	Median
1	173.0	86.5	176.2
2	16.2	8.1	15.2
3	2279.1	1139.6	2414.3
4	210.8	105.4	216.1
5	420.1	210.1	430.1
6	25.7	12.9	25.97
7	102.0	51.0	109.87
total delay	3226.9	1613.5	3452.6

6.5 Implementation Summary

This section provides a brief summary of the steps taken to implement the proposed approach. The actual steps are necessarily very problem specific, but the following will provide general guidance. It is presumed that the objectives of the physical protection system have been established and the goal is to begin characterizing the risks associated with the facility protection system.

The broad goal of this research was the investigation into methods for risk analysis associated with simultaneous, multi-path adversary intrusions into a facility. The new approach discussed in previous sections was developed with the express intent that it could be extended or evolve to account for this broad analysis perspective. However, as noted previously, the research objective was simplified to demonstrate the feasibility of the approach on multiple, independent adversary paths.

6.5.1 Characterize Adversary Paths

The first step in the evaluation of any physical protection systems is to characterize the various adversary paths. This includes, for each possible adversary path, descriptions of possible adversary actions, identification of the delay elements, and the possible detection elements. This is consistent with existing and proposed methods of protection system evaluation. The major difference is in the effectiveness assessment. Given that the adversary paths have been described the following describes the data collection and analysis steps necessary to measure effectiveness.

6.5.2 Collect Historical Data

Prior to conducting any tests, characterize the anticipated uncertainty associated with each of the tasks on the adversary path being evaluated. This analysis may be based on expert judgement, historical testing data, etc. The characterization may be as simple as minimum, median, and maximum delay times. Alternatively, it may be based on physical testing related to similar systems. it is possible, but discouraged, to begin an adversary path analysis in complete ignorance of the times associated with the delay elements along the path. Various methods can be employed to objectively transform this prior information into a probabilistic description of the task times. One proven method is the SHeffield ELicitation Framework (SHELF) [23]. SHELF is a package of documents, templates and software to carry out elicitation of probability distributions for uncertain quantities from a group of experts.

Alternatively, as demonstrated in Section 6.3, even with a minimal amount of historical data it is possible to capture the uncertainty in the task times with sufficient objectivity to avoid unduly influencing the results.

6.5.3 Conduct Testing

The focus then returns to the more traditional component of conducting the new tests. For each task on the path conduct appropriate tests to characterize the delay times associated with taskSince an explicit effort has been made to collect and organize historical data, the prior information should provide insight into areas of focus when setting up test conditions. An additional side benefit of the effort to characterizing historical information is that these data provides some relief on the stress of collecting vast quantities of new test data. New testing can focus on quality of the testing, perhaps investigating more extreme task or barrier situations.

6.5.4 Characterize Adversary Path Risk

In Section 6.4 the historical data, test data, and the final adversary path risk analysis were accomplished in a single stroke. A more insightful approach would have been to first combine the historical and test data and investigate the sensitivities of the relationship between prior and test data. This comparison was implicitly available in Figure 6.4 and Table 6.4, but a more in-depth analysis would have been more appropriate. Intermediate task results that appeared to be too sensitive to prior information would suggest revisiting the accuracy or relevance of the historical data.

The time associated with the adversary delay is now available as an entire uncertainty distribution (Figure 6.5). In contrast to the traditional single point mean values or max-min delay times, this provides considerably more information to the decision maker regarding the facility risk. It is also possible with the current information to bound this CDF with a credibility interval to better understand the true risk.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

The prime assertion in beginning this investigation was that a new delay time calculation methodology was needed in order to better understand realistic system delays beyond the traditional worst case or conservative scenario assumptions. This work has developed a statistically based methodology that utilizes available test and historical data. As believed prior to and confirmed during this project, access delay test data is extremely limited in the amount of data collected for any given attack or breaching method. Previous investigations that used classical statistical methods have generally fallen short of expectations due the inapplicability of frequency based methodologies to these small data sets.

This research has shown that by using Bayesian techniques, mathematically defendable probability density functions representing barrier breaching times can be generated for small data sets. Using these calculated time distributions, a sequential string of tasks/barriers (time distributions) can be combined using convolution techniques to develop a probability density function for the time required to complete the chosen path.

Barrier breaching or task times can be further modified based on human factors data for a given task complexity, environmental condition, or adversary ability. This step represents a corollary to multiplication factors that have been applied in the past for non-barrier delay methods such as visual obscurants. Instead of a simple multiplication of a discrete delay or task time, these factors are applied across the entire range of the individual task or barrier distribution. It was shown that a task multiplication factor of 2 is justified from test data from nuclear power plant high-stress emergency event simulations.

The specific steps needed to conduct the developed Bayesian access delay methodology are as follows:

1. Determine the adversary path to be analyzed. As discussed, this methodology currently only applies to a single chosen path, requiring analyst to discover possible attack paths using traditional security analysis methods.
2. Gather the data for the individual tasks that must be completed or barriers that must be breached for an adversary to successfully complete an attack along the chosen path. This

could require gathering discrete test data points or already calculated probability density functions for the applicable tasks or barriers.

3. If task/barrier probability density functions are not available, compute each PDF using the developed Bayesian method and Markov Chain Monte Carlo analysis.

4. Apply human factors modifiers as necessary based on the task complexity, environmental conditions, and/or adversary skill level.

5. Use a convolution algorithm to compute the PDF for the full path

6. Appropriately use the calculated path PDF for the final delay analysis conclusions such as the probability that the adversary will be interrupted by the response force before the attack.

Because this methodology is new to the delay analysis community, step 6 above will require some additional thought as to how to use the calculated path probability density function in time. It is envisioned that this data could be used in a number of ways:

Calculate a range of delays and probabilities for a given path and set of data. This would be the most basic use of the statistical methodology, and would be similar to the current methodology for access delay analysis. While the calculation of the final delay distribution would be conducted though the developed Bayesian statistical methodology, the analyst could choose the most applicable delay metric: shortest time, median probability time (50% probability), or another probability level of interest (Figure 7.1).

Calculate the risk *if only a certain delay can be achieved.* If a security system analysis determines that the protective force has a fixed response time for whatever reason (funding, physical geography, skill, etc.) the developed Bayesian methodology can provide the data to evaluate the magnitude of adversary success rate. In Figure 7.2, if the time t is the known minimum system response, then it can be seen that percentage of attacks that cannot be interrupted is that below the line. This can also be accomplished as part of the convolution step if a distribution is used to represent the uncertainty in the protective force response time.

Evaluate a required delay based on adversary skill level. As was seen from the test data gathered for this study, it is often unknown (after the fact) what skill level of adversary is represented by that data. If an adversary skill level were to be defined in relation to the percentage of the total delay time distribution, the required delay time to interrupt given adversary could be calculated (Figure 7.3). For instance, if an A level adversary were defined as being able to achieve the top 15% of delay times, the B level adversary defined as achieving delay between say 15% and 60% of delay times, and the C level adversary the bottom 40% of delay times, the PDF is essentially broken into three time ranges of interest. Thus, the facility can choose the required response time to correspond with the adversary in question. How to create this breakdown would need to be developed and agreed upon by the access delay and security communities.

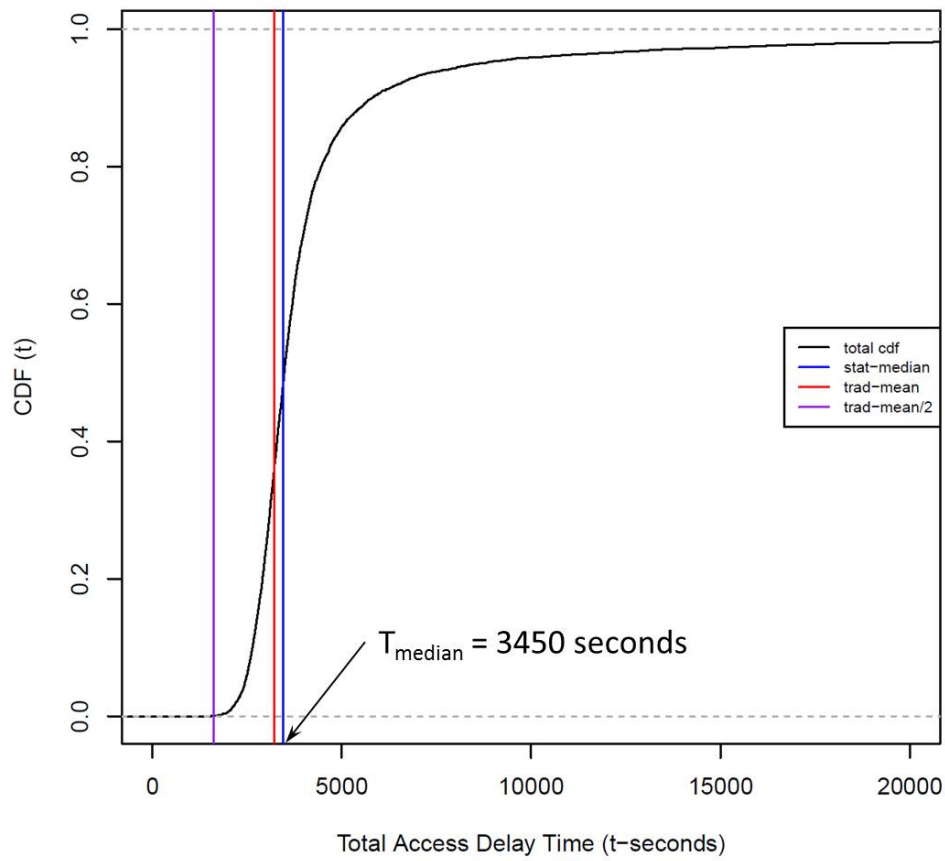

Figure 7.1. Delay as a Function of Median Time

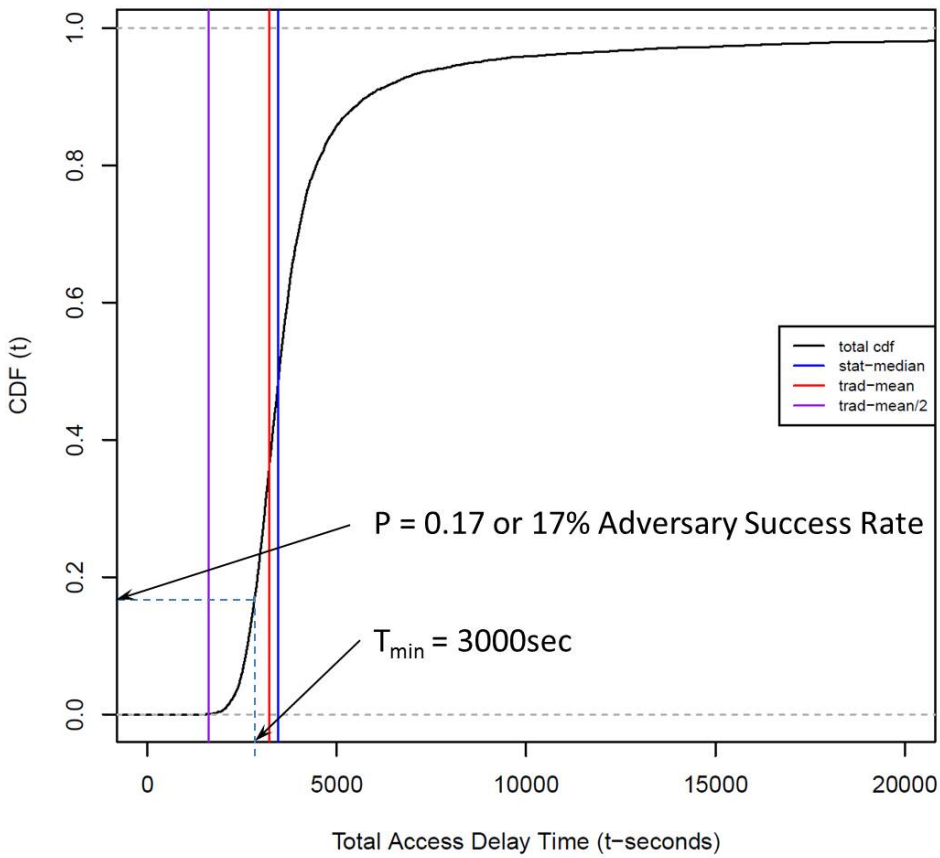

Figure 7.2. Delay as a Function of Minimum System Response

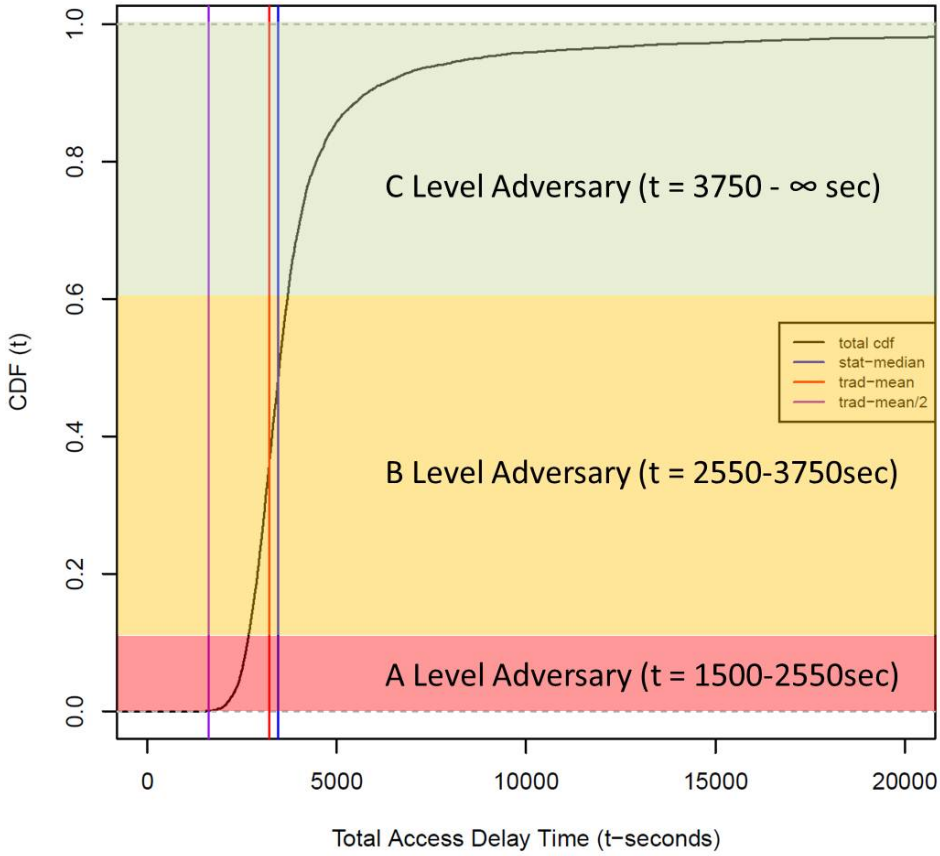

Figure 7.3. Access Delay For Adversary Skill Level

7.2 Future Work

This research has shown that small data sets can be used to calculate probability density functions for tasks and barrier breach times. While one of the goals of this work was to create an easily explained and defendable delay calculation methodology, in its current form the methodology is a mathematical construct and not quite ready for use by delay analysts. The current algorithm implementation was programmed in an open source computing package called *R*, which could be used as an underlying engine with a graphical front end to calculate delay distributions. It is not necessary to use this package, and the method could easily be programmed in a more common computing language like C++. Ultimately the implementation success will rely on an easy to use tool that can call upon a library of delay data as well as allow the user to input new or updated data as necessary. Thus, the next step in the evolution of this technique should be the development of a simple (initially) tool that runs efficiently on a laptop or similar device that the analyst can take into the field.

After a simple working tool has been demonstrated and shown to compute reliable results in a number of situations, the next step would be to incorporate the methodology (or at least its results) into a more complex combat and security simulation tool, such as DANTE. Here, the tool could be used in conjunction with established performance libraries and behavior models to calculate system effectiveness.

While the mathematical basis for the developed calculation technique is believed to be solid, the validity of the techniques output is only as good as the validity of the input data that was used in the calculation. With minimal understanding of how a test or test series was planned, conducted, and reported, no statistical method can adequately predict discrete A, B, or C level performance without significant assumptions being made. If it is assumed that the time delay data is representative of high-, medium-, and low-level adversary skill sets, the methods output will in essence be a distribution covering span of adversary skill sets. While not discretely applicable to A, B, or C level adversary, this combined distribution may still of use to the analyst depending on how that data is used.

Access Delay test data that is being used today has been collected for literally dozens of years, but there is relatively little detailed information on how each specific piece of data was collected or how that data could (or should) compare with other predecessor or successor data. Typically, the objectives for much of the delay testing have been to breach a barrier in the shortest time possible within the limitations of the test itself. These limitations include budget, available test personnel, skill level of available test personnel, tool set considered, tool set available, along with a multitude of other factors that have some potential to influence the final outcome.

Simply, it is difficult to extract adversary level from past reports. Considering there has been no standardized test protocol for conducting delay testing for input into advanced simulations and analysis, the best assumption that can be made about existing reported data is that it was conducted in a manner that best optimized the test conditions for the resources at hand. This would have likely have resulted in an A or B level adversary threat, but even this resolution of skill is unlikely to be recovered. Moving forward, it would be highly advantageous to define a common test protocol for

use in all future delay test activities. At this time, it is difficult to recreate a test series conducted at a different location or by different personnel because there is no standard protocol to which that test or test series was designed or conducted. The development of a standard protocol would be extremely advantageous to the analyst needing to utilize all of the data available to calculate path access delay. An added benefit would be to allow for greater definition of the required resources when proposing and planning a set of access delay tests.

References

[1] VAP-240 ASSESS-ATLAS. URL http://ntc.doe.gov/shared/courses.aspx?section=114. November 2012.

[2] Avert. URL http://www.arescorporation.com/security/products/avert. Last checked: November 2012.

[3] Simajin. URL http://www.rhinocorps.com/RhinoWeb/docs/Simajin_Capabilities_7Apr09.pdf. Last checked: November 2012.

[4] P. L. Ackerman and M. E. Beier. Methods for studying the structure of expertise: Psychometric approaches. In K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, editors, *The Cambridge Handbook of Expertise and Expert Performance*, pages 147–165. New York, NY: Cambridge University Press, 2006.

[5] T. M. Amabile, J. S. Mueller, W. B. Simpson, C. N. Hadley, S. J. Kramer, and L. Fleming. Time pressure and creativity in organizations: A longitudinal field study (working paper). 2002.

[6] H.A. Bennett. The easi approach to physical security evaluation. Technical Report SAND76-0500, Sandia National Laboratories, Albuquerque NM, 1977.

[7] J. Berger and E. William. Choice of hierarchical priors: Admissibility in estimation of normal means. *Annals of Statistics*, 24(3):931–951, 1996.

[8] J. Bernardo. Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society,Series B*, 41(2):113–147, 1979.

[9] G. L. Berry. The weibull distribution as a human performance descriptor. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(7):501–504, 1981.

[10] R. L. Boring, A. M. Whaley, T. Q. Tran, P. H. McCabe, L. G. Blackwood, and R. F. Buell. Guidance on performance shaping factor assignments in spar-h. Technical Report INL/EXT-06-11959, Indaho National Laboratory, 2006.

[11] J. Brewer, S. Hendrickson, S. Cooper, and R. Boring. Human reliability analysis: Informed insights on cask drops. Technical Report NUREG/CR-7016, SAND2010-8463P, Nuclear Regulatory Commission, Office of Nuclear Regulatory Research., 2012.

[12] L. R. Cercarelli and G. A. Ryan. Long distance driving behaviour of western australian drivers. In L. R. Hartley, editor, *Proceedings of the second international conference on fatigue and transportation: Engineering, enforcement and education solutions*, pages 35–45. Canning Bridge, Australia: Promaco Conventions Pty Limited, 1996.

[13] M. T. H. Chi. Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, editors, *The Cambridge Handbook of Expertise and Expert Performance*, pages 21–30. New York, NY: Cambridge University Press, 2006.

[14] S. Dekker. *The Field Guide to Understanding Human Error*. Burlington, VT: Ashgate Publishing Co., 2006.

[15] L Deng and M Wong. An exact algorithm for the statistical shortest path problem. In *Design Automation, 2006. Asia and South Pacific Conference on*. IEEE Press, 2006.

[16] M. L. Garcia. *The Design and Evaluation of Physical Protection Systems*. Butterworth-Heinemann, 2001.

[17] D. Gertman, H. Blackman, J. Marble, J. Byers, L. Haney, and C. Smith. The spar-h human reliability analysis method. Technical Report NUREG/CR-6883, US Nuclear Regulatory Commission, 2005.

[18] Sung Soon Jang, Sung-Woo Kwak, Hosik Yoo, Jung-Soo Kim, and Wan Ki Yoon. Development of a Vulnerability Assessment Code for a Physical Protection System: Systematic Analysis of Physical Protection Effectiveness (Sape). *Nuclear Engineering and Technology*, 41(5):747–752, 2009.

[19] H. Jefferys. *Theory of Probability*. Oxford University, 3rd edition, 1961.

[20] Jouni Kerman. Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics*, 5(0):1450–1470, 2011.

[21] Sandia National Laboratories. DANTE. URL http://umbra.sandia.gov/products.html#dante. Last checked: November 2012.

[22] J. C. Matter. Demonstration of analytic system and software for evaluating safeguards and security (assess). In *30th Annual Meeting of the Institute of Nuclear Materials Management*, July 9-12 1989.

[23] T. O'Haga. Shelf: SHeffield ELicitation Framework. URL http://www.tonyohagan.co.uk/shelf/.

[24] R. Parasuraman and P. A. Hancock. Adaptive control of mental workload. In P.A. Hancock and P.A. Desmond, editors, *Stress, workload, and fatigue*. Mahwah, NJ: Erlbaum, 2001.

[25] M. Plummer. JAGS: Just Another Gibbs Sampler. URL http://www-fis.iarc.fr/~martyn/software/jags/. http://www-fis.iarc.fr/˜ martyn/software/jags/.

[26] I O Ryzhov and W B Powell. Information Collection on a Graph. *Operations Research*, 59 (1):188–201, March 2011.

[27] G Sandine, C Storlie, and S Vander Wiel. Graph Based Statistical Analysis of Network Traffic. 2011.

[28] J. Sharit. Human error. In G. Salvendy, editor, *Handbook of Human Factors and Ergonomics*, pages 708–760. Hoboken, NJ: John Wiley & Sons, 2006.

[29] M.A. Staal. Stress, cognition, and human performance: A literature review and conceptual framework. Technical Report NASA/TM-2004-212824, NASA, Moffett Field, CA: Ames Research Center, 2004.

[30] T. Suzuki, T. L. von Thaden, and W. D. Geibel. Influence of time pressure on aircraft maintenance errors. In *Proceedings of teh 8th International Symposium of the Australian Aviation Psychology Association*, Novotel Brighton Beach, Sydney, 2008.

[31] A. D. Swain and H. E. Guttman. Handbook of human reliability analysis with emphasis on nuclear power plant applications. Technical Report NUREG/CR-1278, US Nuclear Regulatory Commission, Washington, DC, 1983.

[32] C. D. Wickens and J. Hollands. *Engineering psychology and human performance*. Prentice Hall, Saddle River, NJ, 2000.

[33] C. K. Wikle and L.M. Berliner. A bayesian tutorial for data assimilation. *Physica D*, 2006.

[34] R. M. Yerkes and J. D. Dodson. The relative strength of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology*, 18:459–482, 1908.

[35] D. Zakey. The impact of time perception processes on decison making under time stress. In A. J. Maule and O. Svenson, editors, *Time Pressure and Stress in Human Judgment and Decision Making*, pages 59–72. New York, NY: Plenum Press.

Appendix A

Dijkstra Example

As an example, we will populate Figure 4.1 with sample values (Figure D.2) and walk through Dijkstra's algorithm. The results are summarized in Table A.1.

A.1 Dijkstra's Algorithm

A classic algorithm for identifying all the shortest paths in a network with known edge weights is Dijkstra's approach presented in Algorithm A.1.1.

Algorithm A.1.1 Dijkstra(G,c,s)

```
S ← ∅
D_s(s) ← 0
Q ← V
for all u ∈ V such that u ≠ s do
    D_s(u) ← +∞
end for
while Q ≠ ∅ do
    find u ∈ Q such that D_s(u) = min_{u'∈Q}(D_s(u')
    Q ← Q \ u
    for all v ∈ V(u) ∩ Q do
        if D_s(u) + c(u,v) < D_s(v) then
            D_s(v) ← D_s(u) + c(u,v) {The edge (u,v) is selected.}
        end if
    end for
    S ← S ∪ {u}
end while
```

The above analysis is not particularly realistic since we have assumed the most conservative times for intruder breach at each barrier. This generally leads to an overly conservative security system design. The following section discusses an approach that recognizes the uncertainty in the intruder capabilities.

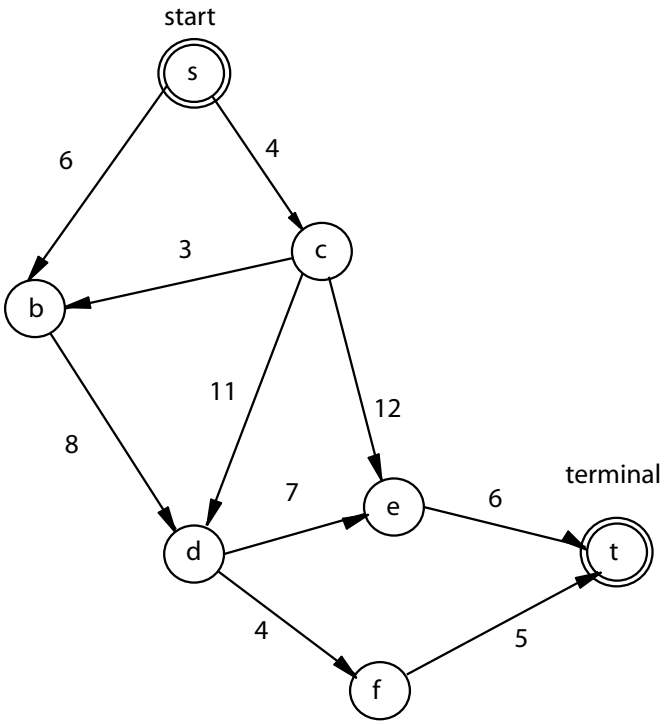

Figure A.1. Populated DAG for Dijkstra Algorithm

Table A.1. Steps in Dijkstra Algorithm

Step	Node	path	Distance to Node from Node 's'	Visited
1	b	{s,b}	6	no
	c	{s,c}	4	no
	d		infinite	no
	e		infinite	no
	f		infinite	no
	t		infinite	no
2	b	{s,b}	6	yes
	c	{s,c}	4	yes
	d	{s,b,d}	14	no
	e	{s,c,e}	16	no
	f		infinite	no
	t		infinite	no
3	b	{s,b}	6	yes
	c	{s,c}	4	yes
	d	{s,b,d}	14	yes
	e	{s,c,e}	16	no
	f	{s,b,d,f}	18	no
	t		infinite	no
4	b	{s,b}	6	yes
	c	{s,c}	4	yes
	d	{s,b,d}	14	yes
	e	{s,c,e}	16	yes
	f	{s,b,d,f}	18	no
	t	{s,c,e,t}	22	yes
5	b	{s,b}	6	yes
	c	{s,c}	4	yes
	d	{s,b,d}	14	yes
	e	{s,c,e}	16	yes
	f	{s,b,d,f}	18	yes
	t	{s,c,e,t}	22	yes

Appendix B

Introduction to Bayes Theorem

B.1 Impact of Testing

Let X be unobservable quantities of interest and D be our data. Then the full probability model can be factored: $P(x,d) = P(x|d)P(d) = P(d|x)P(x)$. Applying Bayes' Rule:

$$P(x|d) = \frac{P(d|x)P(x)}{P(d)} \tag{B.1}$$

The real power of Bayes theorem lies in the ability to combine data in an objective manner. The ability of humans to process probability information is limited at best. The following is a rather classic example of an (important) situation where humans typically do very poorly.

Let's assume that you have tested positive for a disease, breast cancer for example. What is the probability that you actually have the disease? This latter question depends on the accuracy and sensitivity of the test, as well as the (prior) prevalence of the disease.

Assume the following quantities (reasonably accurate numbers):

$$\begin{aligned}
P(Test = positive|Disease = true) &= 0.80 \\
P(Test = negative|Disease = true) &= 0.20\text{(false negative)} \\
P(Test = positive|Disease = false) &= 0.096\text{(false positive)} \\
P(Disease = true) &= 0.01
\end{aligned}$$

Therefore:

$$\begin{aligned}
P(D = true|Test = +) &= \frac{P(T = +|D = true)P(D = true)}{P(T = +|D = true) * P(D = true) + P(T = +|D = false) * P(D = false)} \\
&= \frac{0.80 * 0.01}{(0.80 * 0.01) + (0.096 * 0.99)} = \frac{0.008}{0.103} = 0.078
\end{aligned}$$

So, if a woman has a positive mammogram, there is only a 7.8% probability that she actually has breast cancer. [Only 15% of doctors got this number correct. [ref]] This is a result of the prior assumption that the disease is rare $P(Disease = true) = 0.01$. Alternatively, if you ignore the prior (assume that there is a 50-50 chance someone has the disease), $P(Disease = true) = 0.5$ then:

$$P(D = true|Test = +) = \frac{0.80 * 0.5}{(0.80 * 0.5) + (0.096 * 0.5)} = \frac{0.4}{0.448} = 0.89 \tag{B.2}$$

Of course, this conclusion relies on your belief that 50% of the women in the world have breast cancer. A better approach is to use a plausible prior (e.g. $P(D = true) = 0.01$), but then conduct multiple independent tests; if they all show up positive, then the posterior will increase. For example, if we conduct two (conditionally independent) tests T1, T2 with the same reliability, and they are both positive, we get

$$\begin{aligned} P(D = true|T1 = +, T2 = +) &= \frac{P(T1 = +|D = true) * P(T2 = +|D = true) * P(D = true)}{P(T1 = +, T2 = +)} \\ &= \frac{0.80 * 0.80 * 0.01}{0.80 * 0.80 * 0.01 + 0.096 * 0.096 * 0.99} = \frac{0.32}{0.366} = 0.874 \end{aligned}$$

The probability of breast cancer being present goes from about 8% to almost 90% with the positive results from only one more test. *The value of that one additional test is clear.*

For complicated probabilistic models, computing the normalizing constant $P(d)$ is computationally intractable, either because there are an exponential number of (discrete) values of X to sum over, or because the integral over X cannot be solved in closed form (e.g., if X is a high-dimensional vector). Graphical models can help because they represent the joint probability distribution as a product of local terms, which can sometimes be exploited computationally (e.g., using dynamic programming or Gibbs sampling). Directed graphical models are a natural way to represent many hierarchical Bayesian models.

B.2 Second Example

A second example, slightly more involved, again demonstrates the application of Bayes theorem [**?**]. Consider the toss of a coin where we observe the number of heads, $x \in \{0, 1, \ldots, n\}$, that occur in N tosses. The probability of x heads is characterized by the binomial distribution:

$$p(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \tag{B.3}$$

and the likelihood of generating a sequence of head/tails data is given by:

$$p(d|\theta) = \prod_{i=1}^{n} \theta^{I(x=1)}(1-\theta)^{I(x=0)} \tag{B.4}$$

where the number of heads is given by: $N_1 = \sum_i^n I(x=1)$ and $N_0 = \sum_i^n I(x=0)$. Finally, we note that the constant $p(d) = \int p(d|\theta)p(\theta)d\theta$ can be difficult to compute except in the simplest of cases. The integral is typically solved using Monte Carlo simulation, etc.

B.2.1 Black Swan

Traditional statistics can be used to find a maximum likelihood estimate (MLE) of θ: $\hat{\theta} = \frac{N_1}{N}$.

The difficulty with MLE methods, and traditional statistics in general, is that if we don't see something during test, it is presumed that it can never happen. For example, if we flip a coin 4 times and do not see a head, then the MLE is $\hat{\theta} = \frac{0}{4} = 0$. This is called the *black swan paradox*. This name comes from the fact that for many years of observing swans in England, it was believed that all swans were white and black swans were an impossibility in nature. However, on visiting Australia in 1697, Dutch explorer Willem de Vlamingh discovered that indeed, black swans existed.

B.2.2 Bayesian Approach

B.2.2.1 Prior

Let us assume that the probability of observing a heads on any toss is a random variable. For convenience, let this distribution be a Beta distribution with parameters α, β (α and β are referred to as hyperparameters).

$$P(\theta|\alpha,\beta) \sim Beta(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{B.5}$$

where:

$$mean = \frac{\alpha}{\alpha+\beta} \tag{B.6}$$

$$mode = \frac{\alpha-1}{\alpha+\beta-2} \tag{B.7}$$

$$variance = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{B.8}$$

It remains to identify the hyperparameters α and β. These can be estimated using historical data, expert judgment, etc. For example, to set the hyper parameters of the beta distribution,

suppose your prior is that the probability of heads should be about p, and you believe this prior with strength equivalent to about N samples. Then you just solve the following equations for α and β:

$$p = \frac{\alpha}{\alpha+\beta} \tag{B.9}$$

$$N = \alpha+\beta \tag{B.10}$$

This is just one approach to estimating the prior. We could also have queried experts who might have suggested that the true value of p is in the interval $[p_L, p_U]$ and solved Equations B.6, and B.8 simultaneously for the required hyperparameters.

Caution must be exercised when choosing hyperparameters to assure that the results are consistent. However, in general, the estimates are robust to the choice of the distribution (e.g. Beta) and the associated hyperparameters.

B.2.2.2 Posterior

The posterior distribution can be found by multiplying the likelihood and the prior. Neglecting the normalizing constant for now:

$$p(\theta|d) \propto p(d|\theta)p(\theta) \tag{B.11}$$

$$= [\theta^{N_1}(1-\theta)^{N_0}][\theta^{\alpha-1}(1-\theta)^{\beta-1}\frac{\alpha}{\alpha+\beta} \tag{B.12}$$

$$= Beta(\theta|N_1+\alpha, N_0+\beta) \tag{B.13}$$

It is no coincidence that the posterior distribution is of the same family (Beta) as the prior; our choice of the prior was made with this in mind for this example so we could develop some intuition about the mechanics of Bayesian analysis. The hyperparameters α and β can be thought of as pseudo-heads and pseudo-tails respectively, and $\alpha+\beta$ is the pseudo number of trials assumed prior to the actually testing. This latter value represents the strength of our belief in the prior estimate of p: the larger $\alpha+\beta$, the more strongly we believe that the true value is close to $p = \frac{\alpha}{\alpha+\beta}$.

It is interesting to note that the posterior mean is a weighted combination of the prior mean and the likelihood function. Let $\alpha' = \alpha/(\alpha+\beta)$ and $N' = \alpha+\beta$:

$$
\begin{aligned}
E[\theta|\alpha,\beta,N_1,N_0] &= \frac{\alpha+N_1}{\alpha+N_1+\beta+N_0} && \text{(B.14)}\\
&= \frac{N'\alpha'+N}{N+N'} && \text{(B.15)}\\
&= \frac{N'}{N'+N}\alpha'+\frac{N}{N+N'}\frac{N_1}{N} && \text{(B.16)}\\
&= w\alpha'+(1-w)\frac{N_1}{N} && \text{(B.17)}
\end{aligned}
$$

where $\alpha' = \frac{\alpha}{\alpha+\beta}$ is the prior mean and $\frac{N_1}{N}$ is the MLE mean from the observations. Note that this weighting of the prior information and data likelihood is the result as demonstrated in SectionB.1.

This is also the same concept that is used in modern aircraft navigational systems to update, estimate current flight position and velocity, and predict location at some future time.

Appendix C

Example - Case I: Intruder Detected Immediately

C.1 Background

Case I is a simple example involving eleven sequential tasks. The prior to any testing the distributions for the 11 tasks are assumed to have wide uncertainty:

Tasks 1,3,5,7,9,10 Gaussian

$$\begin{aligned} f(t_i|\mu_i,\tau_i) &= N(\mu_i,\tau_i) \\ \mu_i &\sim N(0,1000) \\ \tau_i &\sim Gamma(0.0001,0.0001) \end{aligned}$$

Tasks 2,4,6,8,11 Gamma

$$\begin{aligned} f(t_j|\alpha_j,\beta_j) &= \Gamma(\alpha_j,\beta_j) \\ \alpha_j,(\beta_j)^{-1} &\sim Gamma(0.0001,0.0001) \end{aligned}$$

The Gaussian distribution is characterized by the mean, μ, and the tolerance $\tau = 1/\sigma^2$. The Gamma distribution is defined:

$$f(t|\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp(-t/\beta) & \text{for } t>0, \alpha>0, \beta>0, \\ 0 & \text{otherwise} \end{cases} \tag{C.1}$$

where: α = shape parameter, β = scale parameter, and $\alpha\beta$, $\alpha\beta^2$ are the mean and variance respectively.

To demonstrate the methodology, sample of task times are drawn from the distributions in Table D.2. For each task, 10 times are generated. In addition, 10 random times for response force reaction are also generated. These samples are used to update the priors described previously.

Table C.1. Case 1: Distributions for Generating Simulated Data

Task	Distribution	Mean	Std Dev	α (shape)	β (scale)
1	Gaussian	0.7000	0.080		
2	Gamma	1.0000	0.410	6.00	0.1660
3	Gaussian	0.1100	0.012		
4	Gamma	0.5500	0.305	3.25	0.1692
5	Gaussian	0.1540	0.018		
6	Gamma	1.0000	0.410	6.05	0.1653
7	Gaussian	0.1100	0.012		
8	Gamma	0.5500	0.305	6.05	0.1653
9	Gaussian	0.1500	0.075		
10	Gaussian	0.1540	0.180		
11	Gamma	4.4785	1.410	10.00	0.4470

C.2 Results

Table C.2. Case 1: Results

Task	2.5%	25%	50%	75%	97.5%
m1	0.5191	0.6281	0.6764	0.7237	0.8285
m2	0.3673	0.8408	1.1466	1.5050	2.4799
m3	0.0793	0.0997	0.1093	0.1185	0.1401
m4	0.1753	0.4371	0.6230	0.8722	1.5760
m5	0.1171	0.1448	0.1583	0.1717	0.2038
m6	0.1374	0.5668	0.9353	1.4216	3.1274
m7	0.0695	0.0956	0.1081	0.1206	0.1511
m8	0.3860	0.8342	1.1035	1.4500	2.3824
m9	0.0277	0.1195	0.1609	0.2032	0.2989
m10	-0.1060	0.1433	0.2509	0.3508	0.5833
m11	2.0306	3.7562	4.8021	5.9972	9.2565
Breach Time	7.0738	9.1341	10.3933	11.9135	15.7362
Response Time	1.9587	2.8399	3.1753	3.4779	4.0515
Delta	3.8615	5.9530	7.2602	8.8323	12.8526

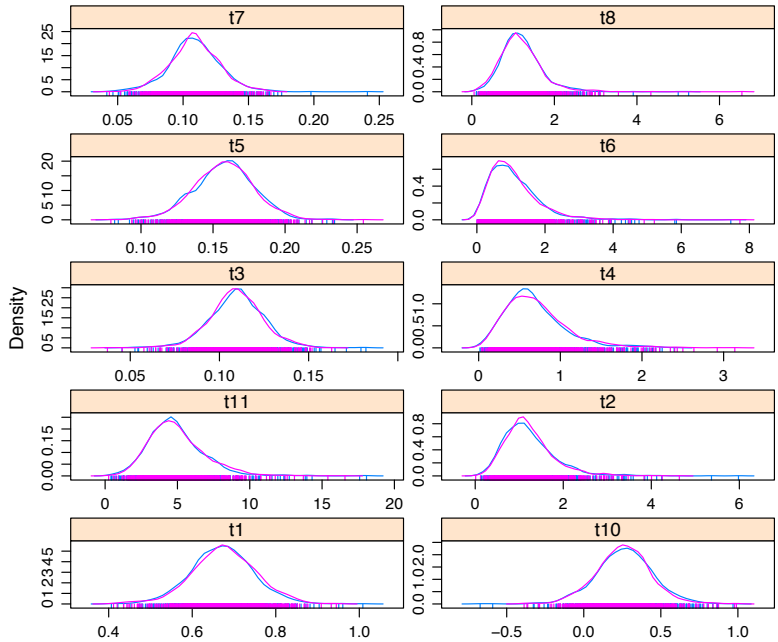

Figure C.1. Case 1: PDF of Task Times

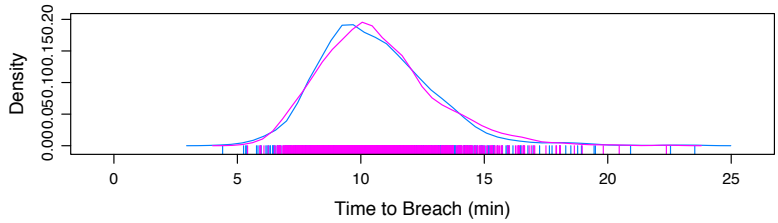

Figure C.2. Case 1: Total Breach Time

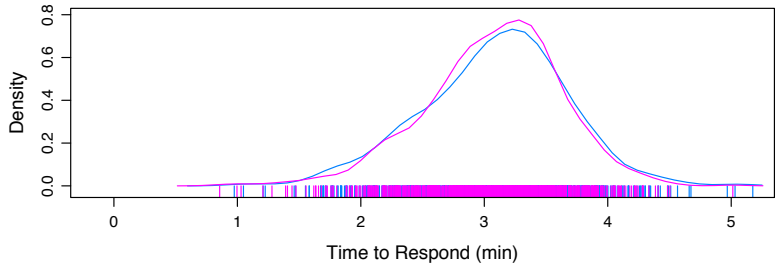

Figure C.3. Case 1: Response Time

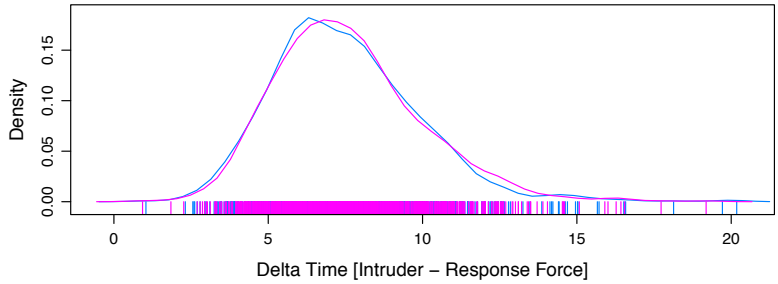

Figure C.4. Case 1: Delta Time (Delta $> 0 =$ success)

Appendix D

Example - Case II: Individual Detection Events

A very simple system will be used to demonstrate the methodology. Consider the system depicted in Figure D.2. The system consists of four barriers where the time to breach for each is a random variable t_i. Further, the probability of detection at each barrier is considered a random variable P_D.

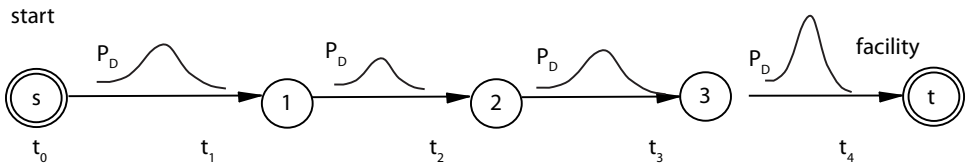

Figure D.1. Demo Problem: Stochastic Breach, Response Times, and Detection Probability

D.1 Uncertainty Foundation

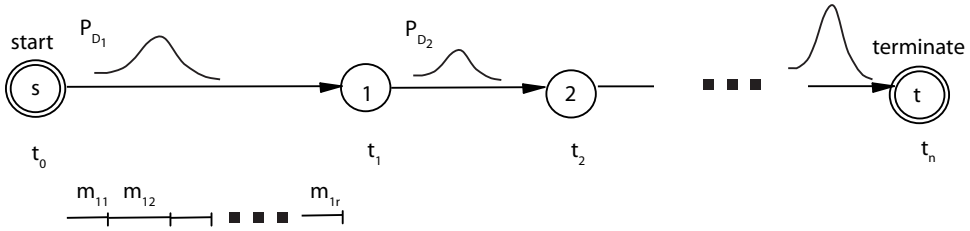

Figure D.2. Demonstration System

The time to breach t will be assumed to be characterized by a Weibull probability density function: *(Note: a gamma distribution would have been appropriate also.)*

$$f(t|\gamma,\alpha) = \begin{cases} \alpha\gamma t^{\alpha-1}\exp\left(-\gamma t^{\alpha}\right) & \text{for } t>0, \alpha>0, \gamma>0, \\ 0 & \text{otherwise} \end{cases} \tag{D.1}$$

with cumulative distribution function: $F(t|\alpha,\gamma) = 1-\exp\left(-\gamma t^{\alpha}\right)$. It is anticipated that it will be convenient to reparameterize the variables by defining: $\lambda = \log(\gamma)$ resulting in:

$$f(t|\lambda,\alpha) = \alpha t^{\alpha-1}\exp\left(\lambda - \exp(\lambda)t^{\alpha}\right)$$

For example, this transformation will allow future consideration for average breach times to be functions of external variables, e.g. defensive measures, intruder exhaustion, etc.

The Weibull distribution is a logical choice since it covers a wide range of possible true distributions and the support is limited to positive breach times, $t_i > 0$.

Further, since very little data is available to characterize the breach times exactly, it will be assumed that true breach time distributions are unknown and therefore the parameters of the distributions are unknown. Specifically, the parameters α and λ independent random variables to be characterized by observational data, expert judgement, etc.

$$\begin{aligned} t_i|\alpha_i,\lambda_i &\sim Weibull(\alpha_i,\lambda_i) \\ \alpha_i &\sim Gamma(a_i,b_i) \\ \lambda_i &\sim N(\mu_i,\tau_i), \quad where \quad \sigma^2 = 1/\tau \end{aligned}$$

Priors will be assigned to α and λ by choosing initial values for $a = 1, b = 0.0001$ and $\mu = 0, \tau = 10^4$ to represent very diffuse intial understanding of the breach times, but operational data will ultimately be used to estimate the distribution parameters of the breach times.

Intruder detection at each point will also be assumed to be unknown, in this case characterized as Bernoulli random variable, d, where the distribution function is defined:

$$f(d|p) = \begin{cases} p & \text{if } d=1 \text{ a intruder is detected,} \\ 1-p & \text{if } d=0 \text{ intruder is undetected} \end{cases} \tag{D.2}$$

and the uncertainty in the detection probability p is described with a Beta probability distribution with parameters θ, η:

$$\pi(p|\theta,\eta) = \frac{\Gamma(\theta)\Gamma(\eta)}{\Gamma(\theta+\eta)} p^{\theta-1}(1-p)^{\eta-1}, \tag{D.3}$$

The Beta distribution is chosen since it is defined over the interval $[0,1]$ and, similar to the Weibull for the breach times, it is extremely flexible in characterizing the uncertainty in p. A very

neutral prior will be assigned to p by choosing initial values for $\theta(=1/3)$ and $\eta(=1/3)$ [20], but operational data will ultimately be used to estimate the distribution parameters and the probability of detection at each point.

D.2 Scenario

For this initial exercise, the initial uncertainty in the probability of detection at each barrier will be assumed to be characterized by a Beta distribution with parameters θ, and η. Since at the beginning of this effort (prior to the consideration of available data) these parameters are assumed to be unknown, this lack of knowledge will be captured with a probability density function on each parameter. The distribution of each parameters will be assumed to be a non-informative Gamma distribution (depicted in Figure D.3).

The Gamma distribution in Figure D.3 indicates that the values of η and θ are more likely to be small, and suggesting that we are assuming a high level of uncertainty in the initial probability of detection. For reference, $\theta = 1$ and $\eta = 1$ results in a non-informative distribution which simply suggests that the probability of detection is equally like to be any value in the interval $[0,1]$; depicted in Figure D.4 as a blue line. As the values for η and θ increase, the uncertainty in the probability of detection decreases suggesting we are more confident in our understanding of P_D. For example, $\theta = 40$ and $\eta = 40$ results in a narrow distribution depicted as a red line in Figure D.4.

The response times of the Security Forces is also assumed to be a Weibull distributed random variable with unknown parameters. The unknown parameters are also assumed to have non-informative prior distributions which are updated with observational data and expert judgement.

The particular prior distributions are chosen to be flexible enough to capture the basic properties of our understanding of the problem, e.g. only positive values, but also general enough to allow observational data to quickly dominate our initial characterizations.

Table D.1. Data Table

t_1	3	1	2	5	3	1	2	5
t_2	10	5	20	40	10	5	20	40
t_3	30	20	40	50	30	20	40	50
t_4	2	2	2	3	2	2	2	3
t_r	11	12	20	10	40	6	5	10
p_{det}	0.9	0.8	0.9	0.9	0.9	0.8	0.9	0.9

For reference, the test data for the breach times t_i and response times t_r, and detection probability p_{det} are summarized in Table D.1. Two chains were run so simulation convergence could be checked (using Gelman's criteria).

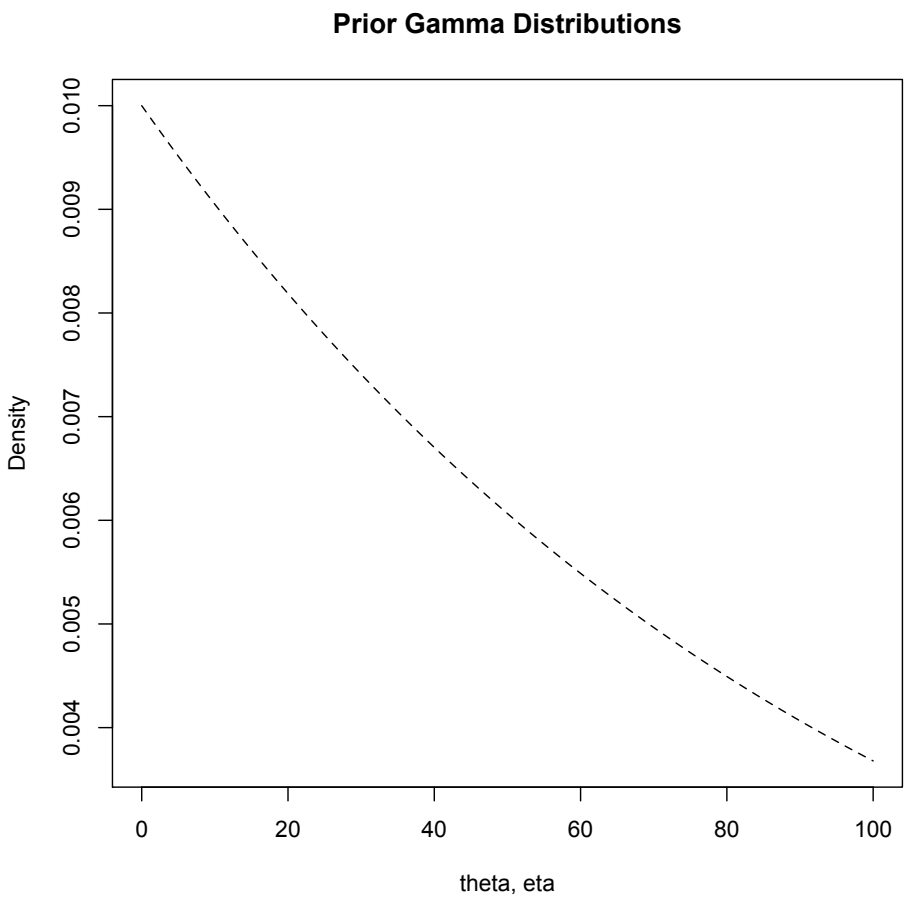

Figure D.3. Prior Distribution on Detection Probability Parameters

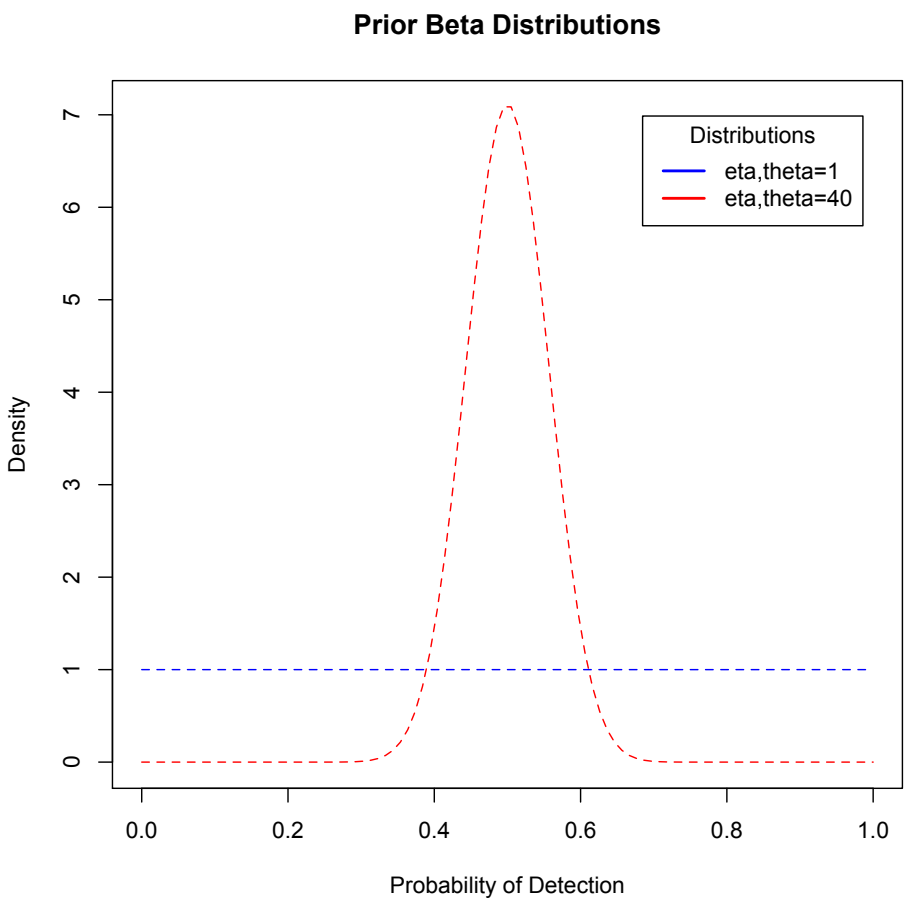

Figure D.4. Prior Distribution on Detection Probability

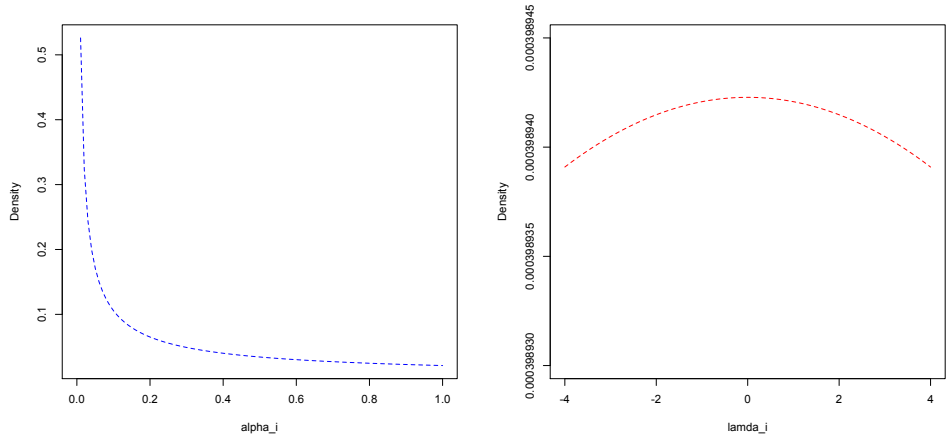

Figure D.5. Prior Distributions on α_i, λ_i

D.2.1 Demonstration Results

Table D.2 summarizes the results of this initial analysis; only the quantiles of representative parameters are provided. Figure D.6 depicts the resulting uncertainty in the probability that the intruders are successfully intercepted, P_s. After consideration of the data related to probability of detection, the resulting uncertainty in P_D is provided in Figure D.7.

Note that there are two superimposed lines on each graph. Each line represents the results from an independent simulation of the problem. It is expected that two simulations (i.e. simulation chains) should result in the similar characteristics and this is visually assessed by superimposing the density functions of the results. In addition, with Markov Chain Monte Carlo convergence is always an issue. Since two chains were run, it is possible to qualitatively check convergence using the Gelman method. Figure D.9 summarizes the results and we see no significant problems (the red line should be close to the black line at the end of the simulation). The variable t_3 should be monitored carefully and perhaps the number of simulations or number of chains could be increased in future analyses.

Table D.2 summarizes the quantiles for the major random variables in the problem. Intermediate variables py_i represent the probability that, given the intruder is first detected at point i, what is the probability that the response team will successfully intercept. For example, looking at the plot of py_1 in Figure D.8, there is a high probability (with high credibility) of intruder intercept at the first breach point (assuming that the intruder was detected). However, if the intruder makes it to breach point 4 before being detected, then we are very confident that the probability of intruder intercept is very low. *Note the short times for breach point 4, particularly relative to the response times (Table D.2).*

	2.5%	25%	50%	75%	97.5%
P_I	0.7301	0.8388	0.8723	0.9015	0.9411
P_D	0.7635	0.8473	0.8792	0.9070	0.9480
t_1	0.5039	1.7976	2.7845	4.0120	7.1111
t_2	1.0118	8.8943	17.1628	27.3798	57.7349
t_3	12.4359	28.3554	36.2884	44.1881	59.8076
t_4	1.0550	1.9408	2.3156	2.6565	3.4026
t_R	0.9916	7.2564	13.6694	21.9656	48.0730

Table D.2. Results From Initial Demonstration Problem

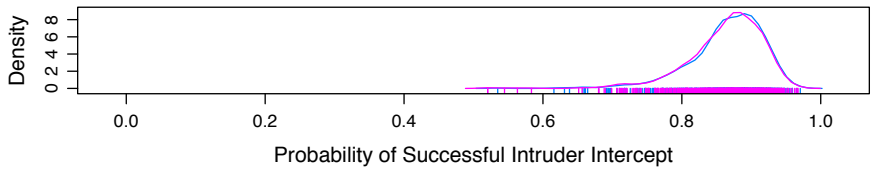

Figure D.6. Probability of Intruder Interception P_s

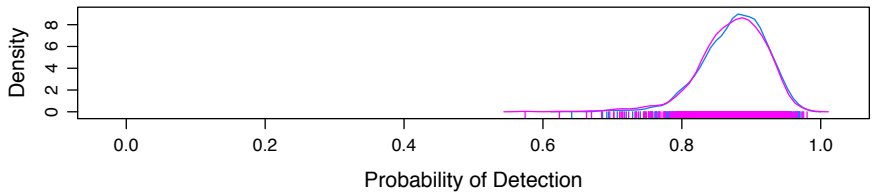

Figure D.7. Probability of Detection P_D

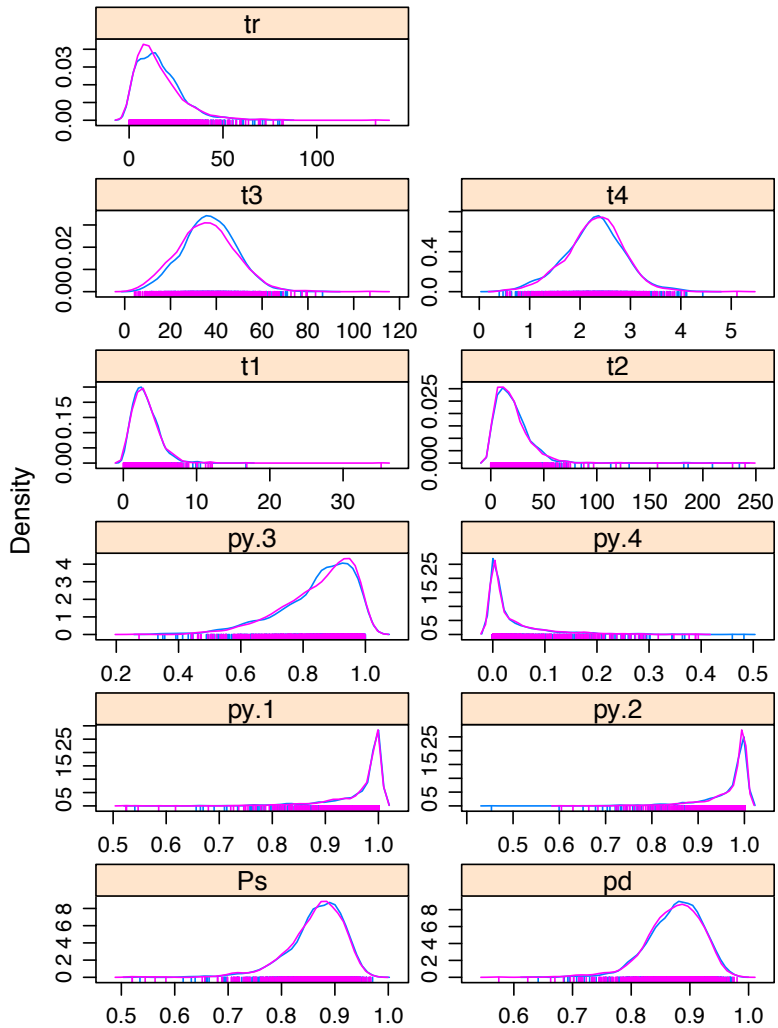

Figure D.8. Summary of Important Random Variables

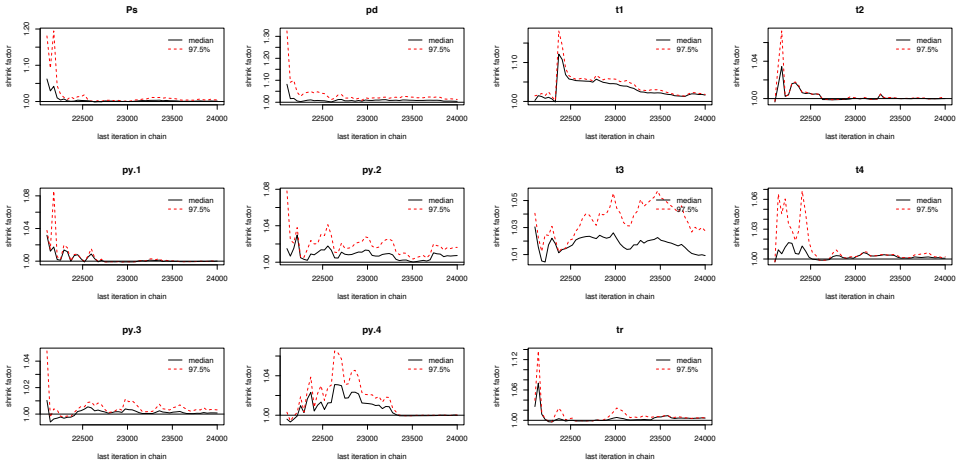

Figure D.9. MCMC Convergence Check - Gelman Statistic

Appendix E

Test Data

Table E.1. Barrier 1 Data

Test	Task 1	Task 2
1	75	113
2	64	106
3	61	100
μ	66.7	106.3
σ	7.37	6.51

Table E.2. Barrier 2 Data

Test	Task 1
1	8.6
2	20
3	20
μ	16.2
σ	6.58

Table E.3. Barrier 3 Data

Test	Task 1	Task 2	Task 3	Task 4	Task 5
1	60	718	1051	292	200
2	100			334	93
3	60				95
4	68				
5	51				
μ	67.8	718	1051	313	129.33
σ	18.98	UNK	UNK	29.7	61.21

Table E.4. Barrier 4 Data

Test	Task 1	Task 2	Task 3	Task 4	Task 5
1	17	29	40	65	69
2	17	24	34	57	70
3	20	27	35	57	71
4	15	28	38	51	77
5	18	26	39	58	72
μ	17.4	26.8	37.2	57.6	71.8
σ	1.82	1.92	2.59	4.98	3.11

Table E.5. Barrier 5 Data

Test	Task 1	Task 2	Task 3	Task 4	Task 5
1	51	86	89	90	114
2	49	72	91	99	113
3	52	74	82	88	117
4	59	67	76	86	133
5	59	59			
μ	54	71.6	84.5	90.75	119.25
σ	4.69	9.91	6.86	5.74	9.32

Table E.6. Barrier 6 Data

Test	Task 1	Task 2	Task 3
1	5	8	14
2	6	9	10
3	6	9	11
4	6	9	12
5	9	10	11
6	6	10	11
7	6	7	13
8	6	7	14
9	6	7	
10	4.5	7	
11	4	6	
12	4	7	
μ	5.71	8	12
σ	1.32	1.35	1.51

Table E.7. Barrier 7 Data

Test	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
1	8	4	5	1	26	18	26
2	8	3	7	1	59		18
3	8	4	15	1			15
4	8	4		1			20
μ	8	3.75	9	1	42.5	18	19.75
σ	0	0.5	5.29	0	23.33	UNK	4.65

DISTRIBUTION:

1 An Address
99 99th street NW
City, State
3 Some Address
and street
City, State
12 Another Address
On a street
City, State
U.S.A.

1	MS 1327	David G. Robinson, 1464
1	MS 1110	Another One, 01400
1	M9999	Someone, 01234
1	MS 0899	Technical Library, 9536 (electronic)

Sandia National Laboratories